UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF NEW YORK

United States of America,

v.

Tyler Scott Johnston,

Defendant.

23-cr-13 (NRM)

MEMORANDUM AND ORDER

For the United States of America:

Margaret Schierberl, Esq.
Tara B. McGrath, Esq.
United States Attorney's Office
Eastern District of New York

For Defendant Tyler Scott Johnston:

Deirdre D. von Dornum, Esq.
Marissa Sherman, Esq.
Austin Dean (Legal Intern)
Federal Defenders of New York

NINA R. MORRISON, United States District Judge:

This is a criminal case brought by the United States against Tyler Johnston. Johnston is accused of committing acts of sexual assault and abuse against his adoptive child ("Jane Doe") between 2019 and 2021, while Doe was approximately 11 to 13 years old. The government contends that the abuse occurred largely at the family home at the Fort Hamilton Army Base in Brooklyn, New York – including on the couch in the living room and on Doe's parents' bed. Also living in the home were

Doe's mother, Monica, and Doe's three younger half-siblings. Johnston denies these allegations, and a jury trial is set to commence on April 1, 2025.

After Doe reported the alleged abuse, law enforcement executed a search warrant on the Johnstons' home, during which they recovered, among other items, the comforter from the Johnstons' marital bed. Law enforcement sent the comforter to the United States Army Criminal Investigations Laboratory ("USACIL"). At USACIL, a DNA analyst sampled six areas of the comforter that, upon further testing, the analyst concluded contained seminal fluid. The analyst then tested the DNA on these stains, concluded that they were DNA mixtures from three or four people. She then compared those DNA profiles to known DNA samples from Doe and her parents, and concluded that Doe and her parents were included as potential donors. Using the probabilistic genotyping software STRmix, the analyst then generated statistics to assess the relative likelihood that these three people, in various combinations, had contributed the DNA in these stains, as opposed to unknown, unrelated persons.

In December 2024, Johnston moved to preclude the government from offering the DNA analyst's testimony at trial under Federal Rules of Evidence 401, 403, and 702. After full briefing, the Court held a *Daubert* hearing that included nearly two full days of testimony, oral argument, and post-hearing submissions.

Johnston's motion to exclude the DNA evidence is granted. The Court does so for three independent reasons. First, the government has failed to demonstrate that USACIL reliably calculated the "likelihood ratios" that accompany the DNA results,

without which the results have no statistical weight or significance. This is because USACIL has not internally validated the available feature of its probabilistic-genotyping software (STRmix) that is applicable to cases where, as here, the defense contends that biological relatives of the persons tested are potential donors of the DNA. Second, the government has not demonstrated the reliability of USACIL's methods for determining the number of contributors in complex DNA mixtures in cases like this one, in which the potential for DNA allele-sharing with biological relatives is high, and in which the analyst was not provided with key information needed to make that determination. For these reasons, the government has not met its burden of showing that the DNA analyst's testimony is reliable evidence that would aid the jury in resolving one or more disputed facts at trial. *See Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993); Fed. R. Evid. 702.

Third, even presuming the reliability of USACIL's testing, all three of the forensic DNA experts who reviewed the data – two for the government, and one for the defense – agree that it is *no more likely* that the DNA profiles seen in these stains were deposited through an act of sexual assault by Johnston against Doe (as the government contends) than through entirely innocent contact with the comforter by Doe and her parents on any number of other occasions (as the defense contends). For this reason, the testimony does not meet even the low bar of relevancy required by Fed. R. Evid. 401; and in any event, its probative value is substantially outweighed by the substantial risk that the proffered testimony would mislead the jury and confuse the issues. *See* Fed. R. Evid. 403.

BACKGROUND

I. Procedural History and Summary of Motion to Preclude DNA

In January 2023, a grand jury in the Eastern District of New York returned a four-count indictment against Tyler Scott Johnston. The government secured a superseding indictment to incorporate what it characterized as non-substantive changes to Count Four on February 28, 2025. This Court conducted four days of jury selection between March 24–27, 2025, and trial is set to begin on April 1, 2025.

All counts in the indictment turn on the government's contention that between December 2019 and April 2021, Mr. Johnston repeatedly sexually assaulted his adopted child,[1] whom the parties here refer to as "Jane Doe," while Doe was between the ages of 11 and 13 years old. Gov. Opp. at 13, ECF No. 122.[2] Johnston pled not guilty and maintains that he did not commit any of the crimes alleged. His counsel has also represented to the Court in various pretrial proceedings that he intends to

[1] Doe is referred to in various exhibits and briefs filed in connection with this motion as Mrs. Johnston's "daughter" and Mr. Johnston's "adoptive daughter." The government has informed the Court that Doe was identified in and outside the family as female at birth and for much of Doe's life. However, some point around or after April 2021, when Doe made the first allegation of sexual abuse against Johnston, Doe began to publicly identify as male and to use "he/him" pronouns outside the home. It is the Court's understanding that Doe continues to identify as male but uses "she/her" pronouns within the Johnston family. After conferring with the parties as to Doe's and the government's preferred references, the Court will endeavor not refer to Doe as "she/her" or as the Johnstons' "daughter" during trial. However, given that all the experts and parties referred to Doe as "she/her" when referencing the DNA evidence, and those terms are also the ones used in DNA analyst's case file, this opinion will use "she/her" pronouns.

[2] Pincites refer to page numbers generated by CM/ECF, and not the document's internal pagination.

assert a fabrication defense – that is, to argue that Jane Doe has made knowingly false allegations of abuse against him – although the contours of that defense will, of course, not be known to the Court and the government until trial begins.

At the time of the events alleged in the indictment, six members of the Johnston family resided together in a military apartment on the Fort Hamilton United States Army base, where defendant Tyler Johnston served as a staff sergeant. The family unit consisted of the defendant; his wife, Monica Johnston; and their four children. The oldest two children – Jane Doe and a son, T.L. – are Monica's biological children, and have different fathers. Def. Mem. at 20, ECF No. 103. In addition, Tyler and Monica have two biological children together, L.J. and M.J., who are Jane Doe (and T.L.'s) younger half-siblings. *Id.* Mr. Johnston adopted Doe, and the couple raised all four children in the same home. *Id.*

In April 2021, when Jane Doe first made these sexual abuse allegations to school officials, Doe was thirteen years old. Doe's half-brother T.L. was twelve years old, and their younger half-siblings were six and three years old. *Id.*

The government contends that over an approximately 17-month period, Johnston coerced and enticed Doe into participating in repeated acts of sexual abuse, which ultimately "escalated" to include vaginal and anal rape. Gov. Opp. at 13. Doe specifically accused Johnston of coercing her into "vaginal sex in his bedroom" – *i.e.,* the bedroom he shares with Doe's biological mother, Monica – on "multiple instances," *id.*, as well as sexual intercourse on a couch in the family's living room. Case Management Notes at 1835A, Def Hrg. Ex. A. Doe alleged that Johnston (who had

undergone a vasectomy) did not use contraception; that he had ejaculated in and outside Doe's vagina; and that on at least one occasion had used his foot to "rub" his seminal fluid into the living room carpet. *See* Gov. Motions in Limine at 40, ECF No. 125; Gov. Opp. to Mot. to Suppress at 8, ECF No. 65. According to Doe, the last act of abuse occurred on April 2, 2021, at which time Johnston vaginally raped Doe on the parents' marital bed while Monica was not at home. Gov. Opp. at 14.

On April 7, 2021, Doe reported this history of alleged abuse to a school guidance counselor, triggering a federal investigation. *Id.;* Def. Mem. At 19. That same night, Special Agents with the U.S. Army Criminal Investigation Division interviewed Doe and her mother in the family's apartment. Def. Mem. at 19. Based on Doe's allegations about where and how the abuse occurred, the agents collected various items from the home. *Id.* at 20. These included three couch cushion covers and cuttings from a carpet in the living room; and bedsheets, a pillow, three pillowcases, and a comforter from the parents' bed. *Id.*

These items were then sent to the United States Army Criminal Investigation Laboratory ("USACIL") for examination and forensic analysis. *Id.* The case was assigned to Forensic Biologist Sara Green, who visually examined the items from the Johnston home with the aid of an alternate light source. *See* Feb. 28, 2025 Hearing Transcript at 19 ("Green Tx."). Green proceeded to conduct serological and DNA testing on certain areas of these household items. *Id.* at 89–97. Green was also provided with DNA reference samples from both parents (Monica and Tyler) and Jane

Doe. *Id.* at 17–18. She was not provided with DNA reference samples from the three other children who lived in the home. *Id.*

On August 30, 2021, USACIL provided the government with a final report authored by Green ("Results of Examination" or "the Report"), which summarized the results of Green's serological and DNA analysis and the primary conclusions and opinions she reached based on her test results. *See* Report at 1695A–98A, Def. Hrg. Ex. A. These findings and opinions included, *inter alia*, that (1) "no semen" was detected on any of the items Green examined, except for the comforter recovered from the Johnstons' marital bed, where she "identified" semen in eight areas; (2) DNA alleles consistent with Monica Johnston's known DNA profile were detected in all eight stains, and the DNA profile in two of those stains "matche[d]" Monica; and (3) six areas of the comforter on which Green concluded that semen was present contained "DNA mixtures" from multiple individuals – five of which, in Green's opinion, contained DNA from three contributors, and one of which contained DNA from four contributors. *Id.* at 1695A–96A.

The Report went on to summarize the results of certain DNA comparisons Green performed. *Id.* at 1697A. In particular, these results concerned the six DNA mixtures she identified on the Johnstons' bedspread (that is, the areas that contained interpretable DNA from more than one donor) against the three known DNA reference samples she had received. *Id.* Using a probabilistic genotyping software called STRmix that is widely used in forensic DNA laboratories, Green calculated the statistical likelihood(s) that the DNA results she obtained from these three- and four-

person mixtures could be explained via certain "competing hypotheses" she had asked the software to calculate. *Id.* The specific hypotheses Green considered, her resulting statistical calculations, and the software and methodologies from which they are derived are discussed at far greater length, *infra*. But in essence, Green's report noted the statistical likelihood that the three-person DNA mixtures in these stains could be explained if the stains included DNA contributions from (1) Monica Johnston, Tyler Johnston, and Jane Doe, as opposed to (2) Monica Johnston, Tyler Johnston, and two unknown, unrelated individuals, or (3) Monica Johnston and two unknown, unrelated individuals.[3] *Id.* Green also tested similar hypotheses, and reported STRMix-generated likelihood ratios, for the four-person DNA mixture she identified on the comforter. *Id.*

Most importantly for purposes of this motion – and the evidence the government now seeks to use at Johnston's criminal trial – Green's report indicated that, in the various scenarios she asked STRMix to consider, the DNA results she obtained were "1.0 quintillion times more likely" to be explained by the presence of Tyler Johnston's and Jane Doe's DNA in the mixtures (along with Monica Johnston's, whose presence was assumed), as opposed to Monica Johnston's DNA and that of two

[3] Although the Report refers simply to "unknown" individuals, at the *Daubert* hearing, Green clarified that the likelihood ratios she calculated using STRMix in fact referred to "unknown, *unrelated*" persons. Green Tx. at 74 (emphasis supplied). This is because – unlike other laboratories – USACIL lacked the capability to use STRMix to consider unknown biological relatives of the individuals in question ("persons of interest") when calculating likelihood ratios. The significance of that limitation features heavily in the defense's *Daubert* challenge, as is discussed *infra*.

unknown, unrelated persons (or, as to the stain with a four-person mixture, three unknown, unrelated persons). *Id.* Similarly, in most of the calculations where Monica Johnston's *and* Tyler Johnston's DNA were "assumed" in the mixture, the DNA results obtained were "1.0 quintillion times more likely" to be explained by the presence of Jane Doe's DNA in the mixtures, as opposed to Monica Johnston's and Tyler Johnston's DNA and one unknown, unrelated person. *Id.*

The government served a Rule 16 expert disclosure on defense counsel on September 10, 2024, along with a copy of Green's report and case file. *See* ECF No. 76. Defense counsel requested an opportunity to speak with Green about her analyses and potential testimony, and Green agreed. On November 19, 2024, two days before her scheduled conference with defense counsel, Green emailed a former USACIL colleague, Prof. Tim Kalafut, listing "FW: relatedness help" as the subject of her query. In the email, Green asked Kalafut, "[S]hould I be concerned for court where I reported semen F1 (probably vasectomized) mixtures of mom/daughter/step-father by assuming mom in the mixtures on her own comforter. Defense is requesting a Daubert hearing on the DNA." 3500-SG-005.

On December 16, 2024, Johnston filed the instant motion to exclude Green's testimony under Federal Rules of Evidence 702 and 403. *See* Def. Mem. The motion seeks to preclude Green's testimony on several grounds, detailed more fully *infra*. Broadly speaking, Johnston's motion challenges both the scientific reliability of the conclusions Green reached, as well as what the defense contends is the ultimate

irrelevance of Green's DNA test results (even if scientifically reliable) to any disputed factual issues in the case that the jury will be asked to resolve.

As to scientific reliability, the motion challenges several aspects of Green's methodology and conclusions. At its heart, the defense's argument turns on its contention that "[t]his case presents a worst-case scenario for forensic biological analysis," in which "several complex mixtures of DNA were collected from a marital bed that could reasonably be expected to include background contributions from six first-order relatives," all of whom lived in the home and had regular contact (unrelated to any act of alleged abuse) with the bedspread. Def. Mem. at 10. Johnston argues that the methodologies available to, and utilized by, Green were grossly inadequate to yield relevant and reliable results in these circumstances, in that, among other things:

(1) Green was not provided with DNA reference samples from three of the four Johnston children who had regular (and in some cases daily) contact with the comforter, despite USACIL laboratory protocols that provided for these samples *and* despite specific requests by Green's case manager that they be sent to the lab;

(2) USACIL had no internally validated protocols to reliably take into account the potentially significant impact of "kinship" when analyzing the number of potential contributors to a DNA mixture and their respective contributions;

(3) Green was provided with biasing information that may have tainted her analysis, further undermining the reliability of her ultimate conclusions;

(4) The STRMix-generated probabilities included in Green's report were unreliable and effectively meaningless here, because USACIL – unlike other labs – had not validated STRMix to account for whether unknown *related* persons (which would include the Johnstons' three other children) might have contributed their DNA to the mixtures in addition to, or instead of, "persons of interest" such as Jane Doe; and

(5) Green's conclusion that the stains on the comforter were "identified" as containing semen overstated the conclusions that could be drawn from the limited testing she performed.

Appended to and cited in the motion were, among other exhibits, an array of studies and scientific literature which, according to the defense, establishes a significant error rate among DNA analysts tasked with interpreting complex mixtures containing DNA from two, three, or more individuals who are related to one another to varying degrees, including studies in which known donors to the mixtures were related to one another to varying degrees. The defense also proffered contrasting examples of other DNA laboratories which, unlike USACIL, had specific protocols and validations that took potential related-family-member contributions to DNA mixtures into account.

As for its Rule 403 challenge, the defense urged this Court find that Green's testimony, even if based on reliable science, has no place at this trial and should be excluded under Federal Rule of Evidence 401 and 403. According to the defense, Green's testimony would not in any way inform or aid the jury's determination of

whether Johnston sexually assaulted Doe on the Johnstons' marital bed (or in any other location), as the government claims and Johnston denies. This is because, the defense argued, it is undisputed that all six members of the Johnston home – including not only the two adults who shared the marital bed, but Jane Doe herself – came into regular contact with the bedspread, and the DNA testing performed by Green does not answer or even bear on the question of *how or when* the DNA from any member of the Johnston family came to be there. Thus, even if the jury were to infer from Green's testimony that DNA from Doe and both her parents were present on areas of the bedspread that also contained Johnston's semen, that means nothing more than that each of these three people had contact with the bedspread at some unknown time(s) before it was sent to USACIL. According to the defense, it is equally likely, as a scientific matter, that Doe's and Johnston's DNA came to be in those locations in any number of ways that involve purely incidental or expected contact – for example, when Johnston had intercourse with his wife or slept in his own bed, or when Doe played with her younger siblings or watched videos on the bed.

In its opposition brief, the government vigorously disputed, first, the defense's claim that Green's results and methodologies were insufficiently reliable to be admitted under Rue 702. The government argued that, *inter alia*, Green used protocols that followed USACIL's rigorous internal validations and external accreditation requirements and were generally accepted in the forensic DNA community, and that Green had accurately and reliably calculated the number of contributors and accompanying statistical weight for all samples. The government

submitted that any quarrel the defense may have with her application of those methods here – whether because of the "relatedness" of potential DNA donors or Green's use of STRmix to generate certain hypotheses – are properly reserved for cross-examination at trial, but are not grounds to preclude her testimony.

The government did not dispute that Green had not received, and thus did not test, DNA standards from fully half of the Johnston household members who had repeated innocent contact with the bedspread. But that fact is of no consequence here, it argued, given Green's reliable determination – based on the evidence she *did* have available – that five stains on the bedspread did not contain DNA from these three untested children. According to the government, Green's results show that these stains were in fact "mixtures originating from three contributors – which profiles matched the defendant, Monica, and Doe[,]" and in which "[t]here were no [DNA] alleles present in Stains 7, 10, 11, 20, [and] 25 that did not originate from those three individuals." Gov. Opp. at 56–57. Indeed, the government argued, the defense's claim that Green failed to account for the likelihood that one or more of the stains she tested also contained DNA from the Johnstons' other biological children, who share all of their parents' DNA alleles, was a mere red herring, since it is a hypothesis that was foreclosed by Green's actual data. *See* Opp. at 66 ("[*I*]*t is not possible* that any such hypothetical DNA could be fully masked here. Moreover, the actual data in this case *makes clear* that there is no additional DNA from these children in the mixtures tested." (emphasis supplied)).

The government appended to its opposition brief an affidavit by a retained consulting expert, Prof. Tim Kalafut – the lead author of one of the "kinship studies" relied upon in the defense's motion, and a former USACIL colleague and "friend" of Green's. Green Tx. at 184. Kalafut opined that the "challenges" that led to the high error rates in the study he authored "are not seen in the DNA mixtures in this case." Kalafut Aff. ¶ 7, ECF No. 122-7. He reached this conclusion because – as he then understood it – the Johnston case did not involve multiple "fully related first order relatives," such as "mixtures of 2 or 3 siblings, or parent/child/child mixtures." *Id.* Kalafut also personally reviewed the data from three of the six DNA mixtures sampled by Green from the comforter. This review gave him confidence that Green had correctly concluded that the Johnstons' untested children were not "masked" (or hidden) contributors in what she believed were all three-person mixtures, because he found absolutely "no indications of biological children masked" in these three stains. *Id.* ¶¶ 47–48.

Kalafut's affidavit acknowledged that he had "not analyzed all relevant stains in this case," but explained that he "selected these stains as representing the overall 'big picture' view." *Id.* at ¶ 47. At the *Daubert* hearing, however, it became clear that Kalafut had not selected this subset of Green's work to review based on his independent assessment that it was representative of the so-called "big picture." Instead, Green acknowledged that she "directed him to the three stains that [she] believed it would be easiest to see[] if there were a masked biological child." Green Tx. at 201.

The Court then held a *Daubert* hearing, which encompassed nearly two full days of testimony followed by argument. On February 28, 2025, Green testified, and on March 7, 2025, Kalafut testified.

Before the hearing, at the government's request, Kalafut conducted a supplemental analysis of the DNA mixture stains that he had not examined before he prepared his January 2025 affidavit. When he did so, Kalafut concluded that Green had, in fact, failed to recognize the presence of what Kalafut concluded was likely a *fifth* contributor to at least one stain on the comforter. Mar. 7, 2025 Hearing Transcript at 53–54 ("Kalafut Tx."). Moreover, Kalafut determined, based on his own review of the data, that the presence of a "masked" biological child of the Johnstons' – *i.e.,* one who shares his parents' DNA alleles – was, in fact, the most likely explanation for the DNA data observed in this four- or five-person stain. *Id.* at 52–53.

After the government's experts testified, the Court heard nearly two hours of oral argument. Although the in-court testimony had, at the Court's directive, focused on the defense's challenge to the reliability of Green's scientific conclusions, the Court gave all counsel the opportunity to address any and all aspects of the defense's motion and the government's opposition at argument. Regarding the defense's motion to exclude under Rule 403 (and the related relevancy prong of Rule 702), the government acknowledged that the DNA evidence did not "prove" the defendant's guilt. *See* March 7, 2025 Oral Argument at 17 ("Oral Arg. Tx."). But it argued that the jury could reasonably infer from Green's DNA analysis that Doe's DNA was "mixed in with

the defendant's semen on his bedspread," and that therefore it is "more likely than not – there is a reason that they mixed . . . and the government submits that the reason is that he assaulted her on that bed." *Id.* at 15. The government also argued that the relative contributions of Doe's DNA in certain stains as compared to the defendant's seminal fluid, a "rich source of DNA," made it more likely than not, as a scientific matter, that her DNA was deposited during one or more acts of rape by Johnston, and thus tended to negate the defense hypothesis that any DNA from Doe was just as likely due to primary or secondary contact that Doe had with her parents' bed on numerous occasions unrelated to any alleged abuse. *Id.* 18. And the government reiterated its view that each of the defense's alternate hypotheses about how Doe's DNA might have been deposited on the bed were properly reserved for argument at trial, but did not justify preclusion of the DNA evidence entirely. *Id.* at 15.

The defense vigorously disputed the government's claims about the "more likely than not" link between Doe's DNA and the alleged abuse on the bed as unsupported by – and contrary to – the testimony of the government's own experts, *i.e.,* that USACIL's DNA results could at most only answer the question of *whose* DNA was found on the bedspread, but provided no information about how, when, and in what sequence(s) the various individuals' DNA came to be there. *See id.* at 28. Permitting Green to present these results the jury, the defense argued, would not even meet the low bar of relevance under Rule 401, since testimony by the government's own experts established that it was not any more likely, as a scientific

matter, that Doe's DNA came to be on the bedspread through an act of rape than through any of the many occasions in which she had incidental contact with her parents' marital bedspread. *See id.* at 32–33. It would also, the defense argued, be misleading to the jury and result in substantial unfair prejudice, since a lay jury would simply do "exactly what the government wants [it] to do" notwithstanding the limitations of Green's results: "put together semen on the bed and Jane Doe's DNA and think it means abuse." *Id.* at 34–35.

To more fully assess the defendant's Rule 403 objection, Court also inquired of counsel for the government what it could proffer about the testimony of its witnesses, particularly Doe, regarding the nature and frequency of her (and her siblings') contact with the bed on other occasions based on its interviews with Doe to date. *See id.* at 22–24. Defense counsel then informed the Court that the Rule 3500 material did not include any information from prior interviews with Doe that addressed this issue in any way. *Id.* at 24. However, the government indicated its understanding that Doe "typically" did not play with her younger siblings on the bed and was not "frequently" on the bed, but was unable to make further proffer at that time. *Id.* at 17, 22.

After the hearing, the government spoke again with Doe and made additional Rule 3500 disclosures. *See* 3500-JD1-50 (Mar. 11, 2025). With leave of the Court, the defense filed a supplemental letter brief citing the new 3500 material on March 19, 2025. *See* Defense March 19 Letter to Court, ECF No. 200 ("Def. 3/19 Ltr."). These new disclosures included statements from an interview with Doe on March 11, 2025, informing the government that in the relevant time period, she recalls spending time

on her parents' bed to watch her tablet (*i.e.*, personal electronic device). *See* 3500-JD1-050. In addition, Doe stated that while the family lived at the Fort Hamilton Army base, the two youngest siblings (the Johnstons' full biological children) slept on their parents' bed *every night*, with the youngest sometimes wetting the bed. *Id.*

The defense argued that these disclosures considerably strengthened its *Daubert* arguments as to (1) the unreliability of Green's analyses, since they increased the likelihood that there were additional DNA contributors to the bedspread stains unrelated to any alleged sexual assault, and underscored the harm caused by Green's inability to account for these scenarios in her testing and statistical calculations, as well as (2) its Rule 403 challenges. *See* Def. 3/19 Ltr. at 1–2. The government responded by letter brief on March 19, 2025, asserting that each of the defendant's claims continued to bear only on the weight, not admissibility, of Green's conclusions. *See* Government Mar. 19, 2025 Letter to Court, ECF No. 202 ("Gov. 3/19 Ltr"). This opinion follows.

II. Forensic DNA Analysis

The fundamentals of forensic deoxyribonucleic acid ("DNA") testing are well established in state and federal court; their fundamental soundness and general acceptance are not in dispute in this case. For necessary context, however, the Court briefly summarizes the methodology used by modern DNA analysts – including USACIL's Forensic Biologist Sara Green – when extracting, analyzing, and assessing the statistical significance of any results obtained from forensic DNA evidence.

A. Overview of Applicable Methodologies

Forensic DNA testing focuses on certain regions, or loci, of DNA that are known to vary between individuals. Green Tx. at 8. The loci that are typed and examined in forensic DNA testing are specific known "repeated DNA sequences" that are "scattered throughout the human genome." John Butler, Fundamentals of Forensic DNA Typing 147 (2010) ("Butler, *Fundamentals*"). Known sequences that are only two to seven base pairs in length are called "short tandem repeats," or "STRs." *Id.* at 148. Examining the differences between the STRs is an effective means of distinguishing between the DNA of different persons, because the number of times any given sequence of DNA repeats at a certain locus is "highly variable among individuals." *Id.*

"The number of repeats at a particular locus constitutes the DNA type or allele present at that location." *United States v. Jones*, 15-cr-152 (VSB), 2018 WL 2684101, at *2 (S.D.N.Y. June 5, 2018). An allele is represented by a number, *see* Green Tx. at 24, and that number is the number of times the known DNA sequence repeats at that particular locus, *see Jones*, 2018 WL 2684101, at *2. Thus, what may vary between individuals is "how many times the piece of DNA is repeated over and over" at a particular STR locus. Green Tx. at 8. At any given locus, a person will possess two alleles, one inherited from their mother and one from their father. *See* Butler *Fundamentals* at 25.

A person's DNA "genotype" is the characterization of their alleles at one or more STR loci. *Id.* For example, if a person inherited an 11 allele from her father

and a 12 from her mother at a certain locus, her genotype at that locus would be "11, 12." *See* Kalafut Aff. ¶ 15. In this case, that person would be known as a "heterozygote" at that locus. *Id.* Alternatively, a person could only have one allele at a locus because they inherited the same allele from both their mother and father, for example, an 11, giving them the genotype "11, 11." *Id.* This person would be known as a "homozygote" at that locus. *Id.* Because a person's genotype at a particular locus is inherited from their mother and father – who each contribute one allele per locus to their child's DNA – a person has only four different possible genotypes at a single locus. For example, if at a certain locus a man has a genotype of "5, 6" and a woman has a genotype of "8, 9," their biological child could have only one of the following potential genotypes at that same locus: "5, 8," "5, 9," "6, 8," or "6, 9." But if those same parents have genotypes of "5, 6" and "5, 9," then their biological child could be a "5, 5" homozygote at that locus (or, alternatively, could have a "5, 9" or "6, 9" genotype).

"A DNA *profile* is the combination of genotypes obtained for multiple loci." Butler, *Fundamentals* at 25. USACIL uses DNA software that examines 21 different STR loci to determine a person's DNA profile. *See* Green Tx. at 9.

At USACIL, as is standard in the field, forensic testing proceeds through four basic steps. *Id.* The first step is extraction. *Id.* As Green explained at the *Daubert* hearing, this process may involve the analyst using a swab to remove biological material from the item of evidence they are testing (in the instant case, a bedspread). The analyst then takes a cutting of the swab and places it into a tube with "chemicals and heat that break[] open the cells and release[] the DNA." *Id.*

The second step is quantitation, where the analyst determines the "total amount of human DNA" in the sample. *Id.* This step permits the analyst to assess whether the amount of human DNA in a sample meets or exceeds the laboratory's minimum thresholds for further DNA analysis.[4]

The third step is amplification. *Id.* Amplification is the process through which the analyst makes additional copies of DNA at certain loci, and at USACIL, "specifically targeting" the 21 loci used to generate a full or partial DNA profile. *Id.* Amplification is done through polymerase chain reaction ("PCR"), which "describes the biochemical process by which the DNA within a sample is amplified, or copied, in order to generate DNA profile results." Sutton Aff. ¶ 9, ECF No. 122-9.

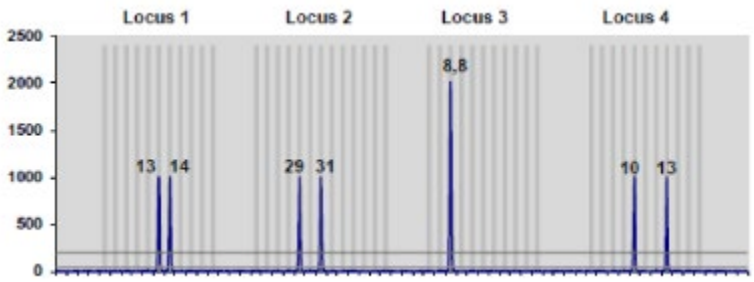
There are some "pitfalls" to the PCR process. Butler, *Fundamentals* at 68. One of these pitfalls is known as "allelic dropout," where one or more of the alleles actually possessed by a DNA donor is not amplified and thus does not appear in the data the analyst examines. *Id.* at 222. Amplification can also create "artifacts," *i.e.,* alleles appearing in the data that are not "real" (because they are not shared by the true donor(s) to a DNA sample), but rather were created in the process of amplification. *Id.* at 217–20.

The fourth step is visualization. Green Tx. at 9. This is done "using a technique called capillary electrophoresis, which is based on the fact that longer

[4] Under USACIL's protocols, the minimum amount of human DNA needed in a sample to perform forensic DNA typing is 25 picograms. Green Tx. at 30. Each of the eight stains tested by Green in this case well exceeded that threshold. *Id.* at 31, 42.

fragments [of DNA] move more slowly than shorter fragments through a polymer solution." *United States v. Lewis*, 442 F. Supp. 3d 1122, 1135 (D. Minn. 2020) (internal citation and quotation marks omitted). It allows "each allele to be distinguished from other alleles." Butler, *Fundamentals* at 175. At USACIL, the instrument used is called "3500xL," which creates an end product that allows the analyst to "visualize the DNA profile" (or profiles) generated from the sample. Green Tx. at 33. The profile appears as a kind of graph, in which each allele is represented as a "peak" at a specific locus. The Y-axis of the graph does not actually represent the number of alleles present, but rather the relative florescence units (RFUs) of that allele, which is a way of measuring the strength of the DNA donor's contribution at that location. Butler, *Fundamentals* at 194.

Below is an "artificial" example of what the visualization may look like from a single source DNA sample – that is, coming from just one person – at four loci when there is "100% interlocus balance between all four loci." John M. Butler, Advanced Topics in Forensic DNA Typing: Interpretation 11 (2014) ("Butler, *Interpretation*"). In this example, at locus three, there is only one allele depicted, but the "peak" is twice as high as the other alleles, indicating that the individual's genotype is "8, 8" at that locus, and the RFU signal at that allele is twice as high. This phenomenon is known as "stacking."



Id. at 11 Fig 1.5.

"Stacking" can also occur when more than one person contributes to a DNA sample. If two or more donors each share the same allele at a particular locus, and the testing yields data for each person's allele at that locus, the allele from each donor will be "stacked one type on top of the other type, and the peak gets higher." Green Tx. at 25.

At USACIL, a software program called Osiris is used to measure the "size of the DNA copies," Green Tx. at 44, as well as to filter out certain "peaks" that may be artifacts, *i.e.*, not "real" alleles from the actual DNA donors' profiles, *see id.* at 80, 195. After that, USACIL analysts take the data generated from Osiris and input it into a different in-house software program called ArmedXpert, *id.* at 189–90, which the analyst uses to "see the DNA profiles," *i.e.*, to "see the peaks" in graph form, *id.* at 46.

At this point, the analyst begins the "analysis and interpretation" step of the process. *Id.* at 10. At USACIL, analysts begin by visually examining the DNA profile data and attempting to determine the "number of contributors" to that profile – that is, the number of different individuals whose DNA might be present in a given evidentiary sample. *Id.* at 48. According to Green, this is done by "looking at how many peaks there are" at a given locus, "looking at the height of the peaks, and then

. . . comparing [them] to the references of the people involved in the case." *Id.* A "single source" profile indicates DNA from just one person. A mixture of DNA profiles may be indicated by the existence of more than two allelic peaks at a single locus. This is because "one person has no more than two peaks at any location," so, for example, a mixture of two people would include "no more than four peaks at any one location." *Id.* at 59. However, some DNA mixtures might contain only two distinct peaks at a single locus if the donors share one or more alleles – for example, if one donor is an 8,8, and the other is an 8,11, the amplified DNA data would, at most, show an "8,11" at that location. *See* Butler, *Interpretation* at 131 (noting that a two-person DNA mixture may contain loci which "have only two alleles.").

Accurately determining the number of contributors to a mixed DNA sample can be difficult and cannot be done by allele count alone. *See* Kalafut Tx. at 99 ("I would never make a number of contributors [determination] based only on maximum allele count . . . ."). That is particularly (but not exclusively) the case when donors are closely related to one another. "Consider a mixed DNA profile containing a mother, a father, and their biological child. Based on allele count alone and barring mutation, the profile can only ever look like it originated from two individuals, since all of the child's alleles are shared with their parents." M.-H. Lin et al., *The Interpretation of Mixed DNA Profiles From Mother, Father, Child Trio*, 44 Forensic Sci. Int'l: Genetics 102175, at 3 (2020). For that reason, analysts typically take into account relative peak height, as disparities in the peak height ratio may indicate the presence of additional donors who share one or more alleles. *See id.* ("A consideration

of peak heights adds more information, and may help inform" the number of contributors determination.). In the case of a DNA mixture from two parents and their child, the child's alleles (at least one of which must be shared by their parents at each locus) would "stack" on their parents' alleles in the visualized data. And if the child is a "minor donor" with a relatively low contribution of DNA in the mixture, that child's profile might "'hide' within the alleles of the other donors." Kalafut Aff. ¶ 16.

Ideally, the peak heights of two alleles at one locus from a single heterozygote donor should be equal, reflecting that person's contribution to the sample. *See* Kalafut Aff. ¶ 29. However, because the amplification process is a biological one, "there is an element of randomness." *Id.* In reality, however, peaks from the same individual might vary in heights (*i.e.,* their reported RFUs) after amplification. According to Kalafut, while there is no hard and fast rule about variation in peak height from a single donor, with taller peaks (peaks more than 2000 RFU), "there may be an expectation of an 80% peak height ratio or better" when looking at two alleles from a heterozygote donor. *Id.* ¶ 30. This means that if "the taller peak is 1000 RFU, the other peak [from that same DNA donor] will be 800 RFU or taller." *Id.*

Where less DNA is present, however, some analysts will consider peak height ratios lower than 80% and still conclude that the allele was contributed by one donor. *See Id.* ¶ 31. In Kalafut's view, a 50% peak height ratio may apply to "lower amounts of input DNA" – meaning that when examining smaller quantities of DNA, analysts

may conclude that two alleles at one locus were contributed by a single donor even if their RFUs vary by as much as 50%. *Id.* ¶ 32.

After estimating the number of contributors, an analyst must "compare the resultant genotype profiles for the possible components of the mixture with the genotypes of reference samples." Butler, *Interpretation* at 145. At this stage, if the DNA profile from a person's known reference sample is consistent with the alleles shown for either a major or minor component of the mixture, "then that person cannot be eliminated as a possible contributor to the mixed stain." *Id.*

USACIL analysts also use a program called STRmix to conduct "probabilistic genotyping" on the DNA profile(s) generated. STRmix is the "market leader in probabilistic genotyping software." *United States v. Gissantaner*, 990 F.3d 457, 466 (6th Cir. 2021) (internal quotation marks omitted). An analyst can use STRmix to "compare reference samples to the deconvoluted evidence samples and provide an overall statistical weight for the comparison in the form of a likelihood ratio." Sutton Aff. ¶ 15.

A likelihood ratio (sometimes referred to as an "LR") does not determine the likelihood that a person's DNA is, or is not, present in an evidentiary sample. *See* Kalafut Tx. at 75. Instead, it evaluates the likelihood of obtaining the evidence results that the analyst has generated under two competing hypotheses: one in which a known person (sometimes referred to as "a person of interest") is a contributor to that evidentiary sample, versus one in which another, unknown person is the source. *See id.* As Green explained, "it compares the idea that . . . this person is included and

their DNA is in the sample, versus the competing idea that . . . it is from some other unknown, unrelated person." Green Tx. at 10–11 (emphasis supplied); *see also* Erin E. Murphy, Inside the Cell; The Dark Side of Forensic DNA 94 (2015) ("A likelihood ratio is a comparison between two different views of the DNA match — one in which the match is explained by the [person[s] of interest] having genuinely contributed to the evidence, and another in which the match is coincidental . . . .") ("Murphy, *Inside the Cell*"). STRmix can also be used, and often is used, to calculate likelihood ratios for various competing scenarios involving multiple DNA contributors – for example, whether Person A and two unknown people are the true donors to the mixture, versus whether the profile came from three unknown people. *See, e.g.*, *Gissantaner*, 990 F.3d at 462 (discussing case in which STRmix was used to compare the likelihood of the composition of a three-person mixture if the defendant was a contributor versus the likelihood of observing that combination of DNA alleles if he was not a contributor).

In an effort to avoid "adventitious" inclusions (*i.e.,* likelihood ratios indicating that people who did not actually contribute DNA are included in the sample), an analyst may "condition[] the LR" on a person's assumed presence in a DNA mixture. *See* Kalafut Aff. ¶¶ 65–70. To condition means the analyst "make[s] an assumption about who donors might be," and has STRmix calculate its likelihood ratio assuming that donor or donors' presence for both competing hypotheses. Kalafut Tx. at 26. For example, in a three-person mixture, an analyst can instruct STRmix to "assume that" a person's DNA profile is in the mixture, and then ask STRmix to calculate the likelihood ratio that applies to other hypothetical scenarios – for example, whether

that "conditioned" donor's DNA is present along with a person of interest and an unknown person, versus one in which the conditioned donor is present with that of two unknown persons. Green Tx. at 173. Conditioning on an individual requires two preconditions: that "the data supports their inclusion," and that "the case scenario" supports their inclusion. Kalafut Tx. at 26.

B. USACIL's Protocols and Validations

USACIL's use of STRmix was "internally validated" in 2014, and it is implemented with "standard operating procedures and in accordance with laboratory accreditation standards and requirements." Sutton Aff. ¶¶ 20, 24. That USACIL is generally accredited in its use of STRmix is not at issue here. Nor is Green's general expertise as a forensic DNA analyst in dispute.

However, the defense's motion relies heavily on what it contends are certain key limitations to USACIL's validations with regard to STRmix generally, and in the methodology Green used. First, USACIL's Technical Leader for the DNA Casework Division Joel Sutton acknowledged that "[a]s it relates to biological relatives, USACIL does not currently evaluate DNA results to distinguish biological relatives using STRmix. Although these calculations can be done using STRmix, USACIL has not internally validated STRmix for this purpose." *Id.* ¶ 24.

Instead, "USACIL uses other mechanisms" to analyze DNA results in cases involving potential donors who are known biological relatives. *Id.* These include "requesting [DNA] references from all these individuals to compare their DNA to the biological evidence in question." *Id.* However, as further discussed *infra,* although

USACIL did request that the government provide reference samples from all six members of the Johnston household, Green only received three.

USACIL's own procedures also warn that "mixtures involving biological relatives that cannot be deduced into individual contributors should be interpreted and reported with caution due to allele sharing." DFSC DNA 114.1 § 6.7, Gov. Hrg. Ex. 226. Additionally, "[m]ixtures determined to be more than four individuals should not be interpreted further due to mixture complexity." *Id.* § 6.6.3.2.

C. USACIL's DNA Testing in Johnston's Case

On or about April 23, 2021, Green received physical evidence in this case. Comforter Chain of Custody, Def. Hrg. Ex. A at 1749A. It consisted of the items collected by the Fort Hamilton CID from the Johnston home (couch cushion covers, and carpet cuttings from the living room, and bedding from the Johnston parents' marital bed), as well as buccal swabs from Tyler Johntson, Monica Johnston, and Jane Doe. Green Tx. at 17–18. Green was told that "trace evidence" was already collected from the comforter, meaning that a "tape grid" had been applied to it to collect "hairs or fibers." *Id.* at 18.

She received the evidence along with a "Forensic Laboratory Examination Request." Lab Req., Def. Hrg. Ex. A at 1783A. The request noted that Jane Doe had reported she had been sexually abused on the bed and the living room couch. *Id.* The request also noted that the Assistant United States Attorney "requests lab examination be expedited to assist with indictment of SSG Johnston." *Id.* at 1785A. Some of the case information sent to Green before she began her testing summarized

the criminal allegations against Johnston as fact, rather than as-yet-unproven accusations of sexual abuse. For example, Green was told that "[Tyler] Johnston has been sexually abusing [Doe] since 2019, incidents would occur on [Tyler] Johnston's bed and living room couch." *Id.* at 1783A.

Before beginning her examination, Green was also informed that Monica Johnston reported to CID that her children also spent time on their parents' bed. Green Tx. at 152. Green's case manager at USACIL, Christine Tarallo, followed up via email to the case agent with a series of questions about the biological relationships among the six members of the Johnston family. Email to SA Abreu, Def. Hrg. Ex. A at 1845A ("Is [Jane Doe] the full biological child of both the subject and Monica Johnston? . . . Do they all have the same parents? Are [Monica and Tyler] the parents of all the other children?"). Tarallo also made the following request:

> Can you get the DNA samples for everyone else that lived in the home? When families are involved, it can complicate our DNA results and having those samples on the front end can help to speed things up when trying to interpret the results and issue the report. I would especially like them considering the mom's statement regarding sex on the couch between her and the subject as well as mentioning that all the kids would be on the bed at some points in time.

Id. Green was aware that this request had been made. *See* Green Tx. at 152. However, Green never received these additional reference samples, and she proceeded to conduct testing and issue her final report without them.

Green began analyzing the comforter with a "focus[] on semen testing." *Id.* at 153. She presumed that Tyler Johnston would be the source of any DNA recovered from semen because "of the three individuals whose references [she] got, he was the

only one who makes semen." *Id.* at 64. The Johnstons' oldest son was twelve years old at the time the evidence was collected, but it appears that Green was unaware of his age and whether he had reached puberty. *See id.* at 153 ("[I]t was not something that occurred to me that there would be someone else in the house that produced semen.").

Green then examined the comforter with an alternate light source, which allowed her to visualize stains that may "fluoresce" under ultraviolet wavelengths and thereby indicate the presence of biological material suitable for DNA testing. *Id.* at 19. Forty-five stains on the comforter "fluoresced" when she did this examination. *Id.* at 92.

Green then proceeded to conduct additional tests on these forty-five areas to determine whether they might contain semen. She began with an acid phosphatase (or "AP") test, and only eight of the stains tested positive. *Id.* She then narrowed her focus to those eight stains and performed a second test on those stains, for Prostate-Specific Antigen, or PSA. *Id.* at 92–93. She described a PSA test as something that "looks very similar to a COVID test," in that there is one end of the testing kit to which liquid is applied, and at the end of ten minutes, a line appears if the test is positive. *Id.* at 92–94. Like a COVID test, the line can appear very quickly or take some time, and a "positive" test can also be very dark or faint. *Id.* at 121–22.

Of these eight stains, Green's notes indicate that seven tested "positive" for PSA. *Id.* at 94. However, she did not photograph the results, did not note how long

it took for the positive indicator line to appear, and did not document how dark or faint the line was. *Id.* at 123.

Acid phosphatase and PSA are presumptive tests for semen. *Id.* at 119. Each test may yield positive results for substances that are not in fact semen. For example, vaginal fluid, blood, and feces can all test positive for acid phosphatase, *id.* at 116–17, and some soaps and urine may produce a positive PSA test, *id.* at 124–25. Green agreed that the only "unique" substance that allows an analyst to confirm that a stain contains semen is the presence of sperm, and she did not visualize sperm in any of these stains. *Id.* at 90, 116. However, Green had been informed that Tyler Johnston was vasectomized; thus, the absence of sperm was not, in her view, inconsistent with the presence of semen here. *Id.* at 95.

Green testified that, even though both AP and PSA are presumptive tests, when she obtains positive results from both on a single stain, USACIL protocols allow her to positively "identify" the substance as semen. *Id.* at 127.

After Green concluded the AP and PSA tests, she discarded the other 33 (semen-negative) swabs from the other fluoresced areas of the comforter. *Id.* at 119. Green did not perform any "substrate control testing" before issuing her final report. That is, she did not test any DNA from the other 33 swabs she took from the other stained areas of the comforter, nor did she test DNA from any non-stained areas, to "understand the background level of DNA" on the item as a whole. *Id.* at 133.

Green then performed a "modified differential extraction" on all eight of the stains that had tested positive for acid phosphatase (stains 7, 10, 11, 13, 19, 20, 23,

and 25). *Id.* at 35. This involved putting that stain's swab in a tube with chemicals and heat to release the DNA. *Id.* After quantitation, Green moved to amplification; she used the program Osiris to determine the "size" of each piece of DNA, and then imported that information into ArmedXpert. *Id.* at 44–46.

Looking at the information in ArmedXpert, Green identified two of the eight stains, 13 and 19, as single source profiles that "matche[d]" Monica Johnston. *Id.* at 50; *see also* Report, Def. Hrg. Ex. A at 1696A ("The F1 DNA profile detected from the comforter . . . stains 13 and 19 matches Monica Johnston"). She explained that even though all of the DNA alleles she visualized for the 21 core STR loci were consistent with a single female donor, this did not change her view that semen was present in these locations. That is because these stains yielded low-level indications of the presence of a minor male donor, but no additional alleles from the core STR loci were detected in the data. *Id.* at 51–52 ("[B]ecause there was so much female DNA compared to just a little bit of male DNA, the female DNA was preferentially copied . . . .").

Green then visually examined the data for the remaining six stains, all of which were DNA mixtures from multiple contributors, to attempt to determine the number of contributors to each. She concluded that Stains 7, 10, 11, 20, and 25 contained DNA from three contributors, and that stain 23 came from four contributors. *Id.* at 62–63. She did this in part by looking at the loci in each stain that contained the largest number of distinct allelic peaks and dividing that number by two. *Id.* at 58–59. For example, stain 11 had no loci with more than five peaks,

so Green's "initial determination of number of contributors for [that] sample [was] three people." *Id.* at 58. Since each person has at most two alleles at a given locus, it would be impossible for any fewer than three people to contribute five alleles. Green agreed, however, that peak heights are also an important factor when evaluating the potential number of contributors. *Id.* at 178. She also acknowledged that the number of contributor assessment had some "uncertainty" when assessing mixtures that could include relatives. *Id.* at 181.

At this point, Green compared the known DNA profiles of Tyler, Monica, and Doe to the six mixture stains. She began by making a determination as to whether Monica was included as a donor, and determined that she was. *Id.* at 63. She did this first because she had already identified Monica as the source of the two single-source stains, and it was "her own bed. . . . so assessing her presence in those mixtures first was the next logical step." *Id.*

She then proceeded to compare the alleles seen in these six mixtures to the DNA profiles she had generated for Tyler Johnston and Jane Doe, *id.* at 64–65, and determined that they were each "included" as potential donors to the mixtures, *id.*

In Green's final report, she concluded only that Monica – but not Tyler and Jane Doe – was a "match[]" for any of the DNA profiles in these stains. *See* Report, Def. Ex. A at 1696A; *see also id.* at 1679A (explaining that the term "match will be used [only] when complete genotypes are obtained at all loci in the amplification kit, and all of the detected alleles are the same as those in the reference being compared").

Green then input the data into STRmix to calculate likelihood ratios for the three- and four-person mixture stains as to various competing hypotheses relating to Monica Johnston, Tyler Johnston, and/or Jane Doe. Green Tx. at 66–67. She directed STRmix to "assum[e]" that Monica, Tyler, and Jane Doe were all "included as contributors." *Id.* at 1695A. Each of the scenarios Green asked STRmix to calculate conditioned Monica as one of the donors to the mixtures; in alternate scenarios, she conditioned both Monica and Tyler as donors. Green Tx. at 67. She directed STRmix to presume Monica's presence each time because she "felt like [Monica] was well represented in the mixtures. . . . [a]nd when dealing with related people, . . . it is best practice to assume one of those people so that they're not impacting the interpretation of the other person." *Id.*

The STRmix likelihood ratios for the specific assumptions and alternate scenarios Green directed the software to consider were reported as follows. The combination of DNA profiles in what Green concluded were three-person DNA mixtures (stains 7, 10, 11, 20, and 25) were deemed "at least 1.0 quintillion times more likely if they originated from Monica Johnston, Tyler Johnston, and [Jane Doe] than if they originated from Monica Johnston and two unknown individuals." Report, Def. Ex. A at 1696A. As for what Green believed was a four-person DNA mixture (stain 23), the resulting profile was "at least 1.0 quintillion times more likely if it originated from Monica Johnston, Tyler Johnston, [Jane Doe], and an unknown individual than if it originated from Monica Johnston and three unknown

individuals." *Id.*[5] She calculated similar "1 in 1.0 quintillion" likelihood ratios for stains 10, 11, 20, and 25 when she directed STRmix to presume both Monica and Tyler as donors and compare the relative likelihood of Jane Doe contributing to the evidence along with Doe's parents, versus the parents and one "unknown" individual. Report, Def. Ex. A at 1697A. The likelihood ratio calculated for stain 7 (presuming the same) was 3.2 quadrillion, and the likelihood ratio calculated for stain 23 was 3.0 quadrillion. *Id.*

Importantly, although Green's report listed the alternate hypotheses above as applying to "unknown individuals," at the *Daubert* hearing, she clarified that any likelihood ratio she asked STRmix to calculate in her USACIL casework only encompassed the likelihood that the DNA profile(s) at issue were contributed by "unknown, *unrelated* individuals." Green Tx. at 10–11 (emphasis supplied). That is because USACIL's current STRmix validations and protocols did not permit Green to consider the likelihood that biological relatives of the known individuals might have contributed some or all of the DNA alleles in these mixtures. *See id.* at 157 (affirming that "USACIL doesn't have validations for untyped or unknown relatives," and that possibility could not be taken into consideration in calculating the likelihood ratios here).

Because Green did not receive reference samples from any of the other children who lived in the house, she did not run any hypotheses that considered the likelihood

[5] This is the largest likelihood ration that STRmix will calculate as "[t]he likelihood ratio is capped at one quintillion." Green Tx. at 83.

of Doe's known half-siblings contributing to the mixtures, either instead of or in addition to Doe. *Id.* Green explained that she was "limited by the references [she] received." *Id.*

At the hearing, Green also testified that when she input the data into STRmix to consider the various propositions she had posed (*i.e.,* one set of conditioned assumptions that presumed Monica's presence, and another that presumed both Tyler's and Monica's), the percentage of DNA in each mixture that the software assigned to each of the assumed three contributors did not change in any significant way. *See id.* at 74–75. Green considered this to be further indication that she "was correct in [her original] interpretation of number of contributors." *Id.* at 75.

At the hearing, Green was also asked about a decision she made to re-run portions of the data in STRmix after her initial run of one stain (#10) initially led STRmix to report that Tyler Johnston was excluded as a contributor. *Id.* at 81–83. Green indicated that the first time she ran her analysis of stain 10, it "failed." *Id.* at 81. When she reexamined the underlying data, she observed that in stain 10 the D1 locus was "missing" a 14.3 allele, which Johnston possesses. *Id.* at 191–93. STRmix did not use the term "failed" (that was Green's characterization), and instead reported this stain as an "exclusion" of Tyler Johnston from the sample, reporting a likelihood ratio of "zero" as to his contributions. *Id.* at 193–94. When Green reviewed the data, however, she concluded that her original interpretation was correct, and that Johnston was not actually excluded as a donor notwithstanding STRmix's report. *Id.* at 77–80. Green testified that in her view, the Osiris program had most likely filtered

out a "real" 14.3 allele of Johnston's. *Id.* at 77. She explained that USACIL's protocols contemplate that erroneous filtering may occur in a "certain percentage of cases," where a contributor's true allele is a single base pair shorter than a second allele that appears at that locus, and that these protocols give an analyst, in certain cases, discretion to recalculate the likelihood ratio to remove the "failed" locus from consideration. *Id.* at 80, 191, 196. Accordingly, Green did not report Stain 10 as an exclusion of Johnston. Instead, she re-ran the data for Stain 10 in STRmix, this time instructing it to "ignore" the D1 locus and rely exclusively on the other loci. *Id.* at 81, 196. USACIL's protocols also permitted Green to go back and re-amplify the DNA swabbed from stain 10 to try and achieve better separation among the alleles at the D1 locus, *id.* at 197, but she did not do so here, *id.*

Green acknowledged that one of the challenges in analyzing DNA mixtures that may contain biological relatives is the risk that a low-level contributor could be present in the sample but "hidden" (or "masked") in the data, due to high allele sharing. *Id.* at 211. She further agreed that an analyst's efforts to avoid "adventitious matches" – an incorrect finding that a person is included as a donor to a mixture, when in fact they are not – is "certainly more complex with biological relatives." *Id.* For example, if a DNA mixture contained contributions from two biological siblings, it might appear as if the mixture included the children's parent (with whom they each share at least half of their alleles) even if the parent was not present. *Id.* at 210.

In this vein, USACIL's own procedures instruct that "mixtures involving biological relatives that cannot be deduced into individual contributors should be interpreted and reported with caution due to increased allele sharing." DFSC DNA 114.1 § 6.7; Green Tx. at 209. The same protocols instruct that for this reason, it may not be possible for a USACIL analyst to provide any statistical weight to a DNA comparison when biological relatives are involved. *Id.*

As for the biological material that might have yielded these results, Green testified that when she swabs a stain, the swab picks up any cells that are in that stain, regardless of its source. *Id.* at 135. Thus, even when DNA is extracted from a stain that tested positive for semen, the profile only indicates the presence of any donor whose DNA is detected in sufficient quantities, "[w]hether they are there from a semen deposit" or another deposit like "skin cells." *Id.*[6]

Green also agreed that her DNA test results provided no information as to how or when any of the DNA profiles from these three or four contributors were deposited onto the comforter. *Id.* at 138. Thus, her testing could not and did not answer the question of whether the various contributors' DNA was deposited on the comforter at the same time; whether some individuals' DNA were deposited at the

[6] One exception is when sperm cells are present in a sample, as sperm cells can be isolated from epithelial cells into separate "F1" and "F2" fractions through a process known as differential extraction. Green Tx. at 26–27, 35. Because there was no sperm in the samples at issue here, however, Green was unable to offer any conclusion or opinion as to what biological material(s) may have produced any of the DNA profiles observed in the mixtures.

same time, and others on a different occasion; or whether they were each deposited at different times. *Id.*

D. Subsequent Review by Government Expert Tim Kalafut

One week after Green testified, the government presented the testimony of a retained expert, Prof. Tim Kalafut. In his original motion, Johnston had attached a paper in which Kalafut was the lead author, citing it for the proposition that DNA mixtures containing related individuals can be difficult to interpret because of allele sharing and may lead to erroneous findings. *See* T. Kalafut et al., *Investigation into the Effect on Mixtures Comprising Related People on Non-Donor Likelihood Ratios and Potential Practises to Mitigate Providing Misleading Opinions*, 59 Forensic Sci. Int'l: Genetics 102691 (2022), ECF No. 103-7 (citing findings by Kalafut et. al. as to high rate of error among DNA analysts in study when asked to determine number of contributors to two- and three-person DNA mixtures containing first-order biological relatives).

Kalafut worked alongside Green as a DNA analyst at USACIL for sixteen years. Kalafut Tx. at 47, 57. He was originally retained by the government to take on the "limited role" of assessing whether the defense properly relied on his research into related-donor DNA mixtures in its *Daubert* motion. Kalafut Tx. at 22, 61. As noted *supra*, he provided the government with an affidavit in opposition to the motion, which was based on his review of a subset of Green's data, after he asked Green to select the samples she believed he should review. *See id.* at 57. After the Court scheduled the *Daubert* hearing, however, Kalafut analyzed the remainder of

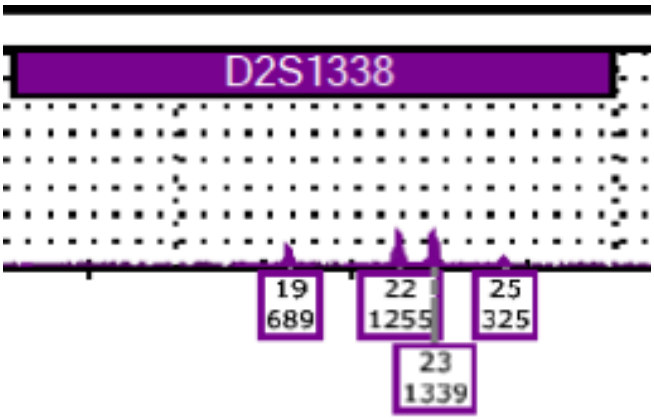
the data from Green's testing. *Id.* at 61. He did not file a supplemental report or affidavit, but at the hearing, he testified about additional conclusions he drew from his review of the complete data.

Kalafut opined that the defense's reliance on his study was a "misunderstanding" of his research. *Id.* at 22–23. His study found a notable 25 to 33% error rate by DNA analysts when determining the number of contributors to mixtures of two or three related persons in a sample. *Id.* at 156. But in Kalafut's view, the Johnston case was different from the scenarios tested in his controlled study of closely related persons, which his team had "purposely designed to be as difficult as it gets." *Id.* at 24. Kalafut based this conclusion on the fact that his study dealt with samples that contained "fully related biological persons," whereas his understanding is that "this case does not involve fully-related people in terms of – in terms of Mr. Johnston, Mr. Johnston, and Jane Doe." *Id.* at 22–23. On cross, however, he agreed that since Johnstons' two biological children also lived in the household, this case could accurately be characterized as involving four "fully related persons" – *i.e.,* those with same degree of relatedness as in his own research – once the relevant framework included all "potential contributors." *Id.* at 74–75. And he agreed that this question – "who are all the potential contributors to this sample"? – is the one that forensic DNA analysts are charged with answering. *Id.*

Kalafut also distinguished his study in that it "focus[ed] on the lowest level [contributors] in an unbalanced mixture where they had different levels of input DNA," whereas in the Johnston case, "the person of interest – that would be Jane Doe

– is not either the trace donor or is not an equal donor." *Id.* at 22–23. However, despite initially referring to Jane Doe as the "clear major donor" in "most" of the mixtures Green tested, *id.* at 23, he acknowledged on cross-examination that Doe was the *minor* contributor in one-third (two out of six) of these mixtures. *Id.* at 157. Kalafut also contrasted his study's finding of high error rates when "balanced" mixtures were at issue (*i.e.,* ones in which the donors contributed roughly equal proportions of DNA to a sample) with the "unbalanced" – and, in his view, much easier to interpret – mixtures in Green's data. *Id.* at 22–24. On cross, however, he agreed that (1) a mixture with two contributors at roughly equal levels, plus a third at a lower (or higher) level, would also qualify as a "balanced mixture," and (2) here, two of the six stains indicated that Monica and Tyler Johnston were the "balanced . . . major contributors." *Id.* at 94–95, 157. Since these two "balanced" donors are also the parents of two biological children in the home, he agreed that at least this portion of Green's work involves "the type of situation [his] study was considering." *Id.* at 158.

Kalafut agreed with Green that there appeared to be three contributors in five of the six mixture stains she tested. *Id.* at 29–31. But he came to a different conclusion than Green with regard to stain 23. Here, Kalafut concluded that there was *likely* a *fifth*, unidentified contributor. *Id* at 53–55. Kalafut specifically pointed the Court to the data from the locus D2S1338, shown below. *Id.* at 50.



ArmedXpert Visualization, Def. Hrg. Ex. A at 1915A. He noted that at this locus, Tyler's alleles are 23, 25, and Monica and Jane Doe both share the genotype of 19, 22. Kalafut Tx. at 50. The substantial difference in peak height between the 23 and 25 alleles (well outside of the lower-bound 50% peak height ratio that would lead an analyst to conclude that they are obligate alleles belonging to the same person) indicated, to Kalafut, that they are likely from an additional, fifth donor. *Id.* at 53–54. And he further concluded that a "masked" biological child of the Johnstons' was the most likely source of this additional contribution. *Id.*

Kalafut also testified that the data indicated to him the potential presence of a fourth contributor in a different stain (#7). Ultimately, Kalafut did not opine that the stain came from four donors and agreed with Green that it was "most appropriate" to characterize stain 7 as coming from three donors. But his testimony was hardly unequivocal on this point. *See id.* at 116 ("There is uncertainty here. [But] my vote is three.").

Kalafut further agreed with a series of foundational propositions that are closely tied to the defendant's *Daubert* challenges. First, although he continued to point to what he believed were key differences between the propositions tested in the

study and the instant case, he agreed that researchers have, in various blind studies, documented a higher error rate when analysts are charged with estimating the number of contributors in mixtures that involve biological relatives. *Id.* at 79. He also acknowledged that even when two potentially obligate alleles from included donors are within a 50% peak height ratio, additional contributors could still be masked. *Id.* at 92.

Kalafut also confirmed that a likelihood ratio does not estimate the likelihood that a particular person is the donor, but simply gives the probability of competing hypotheses from the observed evidence. *Id.* at 124. Even a high likelihood ratio does not necessarily reflect the "ground truth" of whose DNA is present in the sample, and "the best explanation for the data" may be a different, "third proposition" that an analyst has not yet run. *Id.* at 125. Kalafut further agreed that when an analyst chooses various hypotheses to run in STRmix, she should endeavor to specifically test any and all case-specific defense hypotheses – that is, "the defense should be entitled to all alternatives consistent with exoneration" when an analyst uses probabilistic genotyping to calculate a likelihood ratio. *Id.* at 126.

According to Kalafut, analysts should also use STRmix to test defense hypotheses involving untested biological relatives. He himself has written about this issue, noting that when the appropriate features of probabilistic genotyping software are enabled, they allow an analyst to calculate a likelihood ratio based on propositions such as, for example, whether "a person of interest's brother [who is] unavailable for DNA analysis is the source of the DNA." *Id.* at 126, 128. And he confirmed that

STRmix has the ability to run likelihood ratios in precisely these kinds of scenarios, which an analyst simply designates as an "unknown related" hypothesis. *See id.* at 128 ("Q: If you're putting it into STRmix, that would be called an unknown related, right?" "A: Yes, yes"). But that option was not available to Green when she ran the likelihood ratios on the data from the Johnstons' bed:

> Q: So, here USACIL did not run a proposition that allowed for an unknown or untyped related contributor, right?
>
> A: That's correct.
>
> Q: And that's because USACIL didn't have that capability, right?
>
> A: They don't have that function of STRmix turned on.

Id. at 129. By way of example, he agreed that these limitations prevented Green from testing at least one specific defense hypothesis: the likelihood that the mixture was comprised of DNA from Jane Doe, Monica, and one of Monica's other children (which the software would consider to be an "unknown related person"), rather than from Doe, Monica, and Tyler Johnston. *Id.* at 130.

Lastly, Kalafut agreed that "DNA results on their own have no intrinsic meaning without context." *Id.* at 131. For that reason, it is "important to consider what question is actually being asked or answered by a specific forensic test or method" when considering what data is "relevant" to a particular proposition. *Id.* Here, he agreed, the testing Green performed was addressed only to the question "of whose DNA it might be" on the comforter. *Id.* at 132. On the other hand, Kalafut affirmed, Green's testing did not in any way "address[] the question of how the DNA may have been deposited" on the bedspread. *Id.*

LEGAL STANDARD

Federal Rule of Evidence 702 allows admission of an expert witness's testimony:

> if the proponent demonstrates to the court that it is more likely than not that: (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue; (b) the testimony is based on sufficient facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert's opinion reflects a reliable application of the principles and methods to the facts of the case.

Fed. R. Evid. 702. In deciding whether to allow expert testimony, the Court must find that the expert's testimony "both rests on a reliable foundation and is relevant to the task at hand." *United States v. Williams*, 506 F.3d 151, 160 (2d Cir. 2007) (quoting *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 570, 597 (1993)). The burden rests on the proponent of the evidence – here, the government – to establish its admissibility. *Id.*

In *Daubert*, the Supreme Court "enumerated a list of factors that, while not constituting a 'definitive checklist or a test,' a district court might consider in evaluating whether a proffered expert opinion has the required indicia of scientific reliability" under Rule 702. *Nimely v. City of New York*, 414 F.3d 381, 396 (2d Cir. 2005) (quoting *Daubert*, 509 U.S. at 593–94). Those factors include whether (1) a theory or technique can be and has been tested; (2) it has been subjected to peer review and publication; (3) there is a high known or potential rate of error and there are standards controlling the technique's operation; and (4) the theory or technique

enjoys general acceptance within a relevant scientific community. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 149–50 (1999) (citing *Daubert,* 509 U.S. at 592–94).

The *Daubert* factors, however, are not "necessarily nor exclusively appl[ied] to all experts in every case." *Williams*, 506 F.3d at 160 (quoting *Kumho Tire*, 562 U.S. at 141). This is because "the district court's inquiry into the reliability of expert testimony under Rule 702 is a flexible one." *Id.* (internal quotation marks omitted). Thus, "the law grants a district court the same broad latitude when it decides *how* to determine reliability as it enjoys in respect to its ultimate reliability determination." *Id.* (quoting *Kumho Tire*, 562 U.S. at 142).

Challenges to specific methodologies or the strength of an expert's credentials ordinarily "go to the weight, not the admissibility" of the expert's testimony. *McCullock v. H.B. Fuller Co.*, 61 F.3d 1038, 1044 (2d Cir. 1995). However, the Circuit has also clarified that "when an expert opinion is based on data, a methodology, or studies that are simply inadequate to support the conclusions reached, *Daubert* and Rule 702 mandate the exclusion" of the expert's testimony. *Ruggiero v. Warner-Lambert Co.*, 424 F.3d 249, 255 (2d Cir. 2005) (quoting *Amorgianos v. Nat'l R.R. Passenger Corp.*, 303 F.3d 256, 266 (2d Cir. 2002)). The Court must remain vigilant in its role as "the ultimate gatekeeper," ensuring that the proponent of scientific testimony has demonstrated its reliability before it may be presented to a jury. *Williams*, 506 F.3d at 160 (internal quotation marks omitted).

In addition to determining that the expert testimony is based on a reliable application of reliable principles and methods, Rule 702 demands the district court

assess whether an expert's testimony "will help the trier of fact to understand the evidence or to determine a fact in issue." Fed. R. Evid. 702. The *Daubert* Court described this aspect of Rule 702 as "fit," and warned that "'[f]it' is not always obvious, and *scientific validity for one purpose is not necessarily scientific validity for other, unrelated purposes*." *Daubert*, 509 U.S. at 591 (emphasis supplied).

"[E]xpert testimony is [also] subject to Rule 403, and 'may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury.'" *Nimely*, 414 F.3d at 397 (quoting Fed. R. Evid. 403).

The instant motion involves a case-specific challenge to expert testimony on forensic DNA analysis. With over thirty years of broad acceptance in state and federal courts, forensic DNA analysis has become – with good reason – perhaps the most trusted and discerning form of scientific evidence that the legal system has ever known. But that does not absolve a trial court of its gatekeeping function whenever a party seeks to call even a well-qualified DNA analyst as a witness. As the New York Court of Appeals has emphasized:

> In the criminal justice system, [genetic biology] has provided forensic science with one of the most powerful tools for identification yet seen. DNA testing has become the 'gold standard' of this process. For this reason, more than any other, courts must use the tools available to make sure the highest standards of reliability are maintained.

People v. Williams, 147 N.E.3d 1131, 1134 (N.Y. 2020).

In criminal cases, the prejudice that can result from the admission of DNA testimony whose foundation is unreliable, or which is not relevant to the issues in

dispute before the jury, has been well understood for some time. For "[t]he vast majority of jurors have no independent ability to interpret the probative value of results" and therefore "[t]he potential prejudicial impact is unusually high, because jurors are likely to overestimate the probative value of a 'match' between samples." Executive Office of the President, President's Council of Advisors on Science and Technology ("PCAST"), Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods, 45 (Sept. 2016), available at www.justice.gov/d9/2023-07/07.13.23.%20-%20PCAST%20-%20Interim.pdf ("PCAST Report").

It is also true that subjectivity pervades "even the most favorable conditions of forensic DNA typing." Erin E. Murphy, *The Art in the Science of DNA: A Layperson's Guide to the Subjectivity Inherent in Forensic DNA Typing*, 58 Emory L. J. 489, 509 (2008). Those concerns are particularly pronounced when the evidence offered is a complex mixture from multiple individuals, who may share any number of DNA alleles – either by chance, or because they are related to one another. *See* PCAST Report at 8 (discussing the "inherently difficult" process of performing "DNA analysis of complex mixtures").

In the context of a probabilistic genotyping software program, "[r]esults cannot be simply categorized as 'reliable' or 'unreliable' without considering context. In addition, reliability cannot be established *without validation tests* using known samples of similar complexity." John M. Butler, et al., *DNA Mixture Interpretation: A NIST Scientific Foundation Review*, at 5 (2024), available at

https://nvlpubs.nist.gov/nistpubs/ir/2024/NIST.IR.8351.pdf ("*Foundation*") (emphasis supplied).

Thus, in determining whether a challenged use of DNA evidence is sufficiently valid and reliable, a court must not only look to the general reliability of DNA analysis or statistical tool that gives meaning to any potential inclusion of a known individual in a mixture (here, probabilistic genotyping). It is just as critical to ask whether the specific application of the methodology, software, or protocol in the case at bar has been subjected to appropriate validation by the analyst and her laboratory. This inquiry is not just a fundamental tenet of DNA science. It is directly linked to *Daubert's* requirement that courts assess not only whether a given form of scientific evidence is sound, but whether, in a particular case, the "[p]roposed testimony [is] supported by appropriate validation." *Daubert*, 509 U.S. at 590

DISCUSSION

Johnston's motion broadly challenges both the reliability and relevancy of the DNA evidence the government seeks to offer. He makes five core arguments in support of his motion to preclude: (1) because USACIL did not have internally validated procedures to take kinship into account when using STRmix to generate likelihood ratios, Green's conclusions are both unreliable and will not aid the jury's resolution of any factual dispute at trial; (2) Green's conclusions about the number of DNA contributors on the comforter are unreliable; (3) Green's conclusions that the stains she analyzed contained semen are unreliable; (4) Green's analysis was adversely impacted by biasing information she received at the outset of the

investigation, which led her to "s[ee] her assignment as looking at the evidence to see if it matched [the] allegations," Def. Reply Mem. at 15, ECF No. 135; and (5) even assuming, *arguendo*, the scientific validity of Green's test results, the presence of Doe's and Johnston's DNA on this common household item is irrelevant to the jury's determination of whether Johnston sexually assaulted Doe and should be precluded under Rules 401 and 403.

For the reasons to follow, the Court agrees with Johnston's first, second, and fifth arguments. Namely, the Court finds that (1) USACIL's lack of internally validated procedures to utilize STRmix to consider the presence of unknown, *related* individuals renders Green's proposed testimony unreliable for the purpose for which it is being presented; (2) the methodology Green used to determine the number of contributors to these DNA mixtures was not reliable because USACIL had no validated protocols for approaching the kinds of challenges presented here, and was further undermined by the lack of key information provided to Green in this case; and (3) even if all of Green's testimony was based in reliable science, her proffered testimony is of little to no relevance to any fact in dispute at trial, and would, in any event, be substantially outweighed by the danger of misleading the jury and confusing the issues.

I. USACIL Lacked the Necessary Internal Validation to Calculate Relevant and Admissible Likelihood Ratios Using STRmix

The first problem with the government's bid to admit Green's DNA testimony is that it has failed to show that the statistics she would present to the jury to give meaning to her test results – that is, the "likelihood ratios" she generated for the

various DNA mixtures she tested – "rest on a reliable foundation" and are "relevant to the task at hand." *Williams*, 506 F.3d at 160 (quoting *Daubert*, 509 U.S. at 597). That is because USACIL has not validated STRmix for the specific feature in the software that allows an analyst to consider "unknown, related" persons in its calculations. That feature is the only way a DNA analyst can assess the relative likelihood that contributions to a particular DNA mixture came from specific "persons of interest" (POIs), as opposed to one or more persons whose DNA was not tested *but who are biological relatives* of the POIs. And in light of what both parties have made clear are the core factual disputes that the jury will be asked to resolve, that is the only "competing hypothesis" about the DNA evidence that actually matters at Johnston's trial.

Because USACIL had not validated this available feature of STRmix, Green had no option to utilize it. As such, she was left with "1 in 1.0 quintillion" likelihood ratios that sound highly discerning, but in fact are meaningless to the resolution of any factual dispute that will be presented to Johnston's jury about the potential source(s) of the DNA evidence. For they say nothing about the relative likelihood that the DNA profiles seen in these comforter stains were deposited by Jane Doe, her mother, and the parent (Johnston) whom Doe claims sexually assaulted her on that bed (*i.e.,* the "POIs")), as opposed to the likelihood that they were deposited by any other three- or four-person combination of the six members of the Johnston household, all of whom are biologically related to at least one (and in some cases, four or five) of the others.

A. STRmix Functionality, Likelihood Ratios and Internal Validation

Before turning to the limitations of USACILs capabilities and its implications, the Court sets forth a few related foundational principles of DNA analysis. None of these are in dispute here. But they are essential to understanding the testing Green conducted on the Johnstons' comforter in 2021.

First, where a DNA analyst concludes – as Green did here – that one or more known individuals are included as potential donors to an evidentiary sample, that inclusion has zero significance without an accompanying statistical "weight." Green readily confirmed as much at the *Daubert* hearing, *see* Green Tx. at 10 ("when I look at DNA evidence and I determine that someone's DNA could be included, I have to give a weight to what that means"), and it has been similarly recognized by the courts for decades. *See, e.g., United States v. Porter*, 618 A.2d 629, 640 (D.C. 1992) ("[A] match between two DNA samples means little without data on probability").

Second, the vehicle that most forensic DNA laboratories currently use to give weight to potential inclusion of known donors in a DNA mixture – probabilistic genotyping ("PG") – enjoys broad acceptance in state and federal court. The general acceptance of PG software to calculate likelihood ratios when a laboratory has properly validated it for use includes, but is not limited to, STRmix itself. *See, e.g. United States v. Green*, 2024 WL 4488577 (2d Cir. Oct. 15, 2024) (summary order) (finding no error in a district court's decision to admit STRmix evidence without holding a *Daubert hearing* as it "is widely used in forensic laboratories across the country") (internal quotation marks omitted); *United States v. Tucker*, No. 18-cr-119

(SJ), 2020 WL 93951, at *4 (E.D.N.Y. Jan. 8, 2020) (noting that "[c]ourts have overwhelmingly admitted expert testimony based on STRmix results" and collecting cases).

Third, whether a probabilistic genotyping program has achieved wide acceptance and found to pass muster under *Daubert* as a general matter does not end the inquiry as to its reliability in every case. As with all DNA protocols and methods, a laboratory must *internally validate* these PG program(s) for use in its own casework. Internal validation is a rigorous process that measures the analysts' proficiency and accuracy in using the methodology in-house, under conditions that strive to replicate the lab's actual casework. *See, e.g., Foundation* at 65–66, ("Forensic laboratories conduct internal validation experiments before implementing a new technique to assess method performance under specific conditions . . . . Auditors as part of an accreditation review examine validation and look for the types of experiments conducted as part of their approval process." (footnote omitted)); *id.* at 15 (discussing need for internal validation of probabilistic genotyping software and other DNA technology and emphasizing that "reliability cannot be established without validation tests using known samples of similar complexity").

For this reason, when a DNA analyst is asked to testify about DNA results or statistical propositions that exceed the bounds of her laboratory's then-existing validation, courts have appropriately exercised their gatekeeping function to exclude it. Recently, for example, the district court in *United States v. Ortiz*, 736 F. Supp. 3d 895 (S.D. Cal. 2024), granted a motion to preclude the introduction of testimony in

which STRmix had been used to interpret a DNA mixture swabbed from a firearm, where the record indicated that the sample may have contained as many as six contributors. The court acknowledged the foundational validity of STRmix itself, but found that the particular application at issue was precluded by Rule 702 and *Daubert* because, *inter alia,* "STRmix had not been subjected to developmental validation for six-person mixtures by the developer or internal validation by the SDPDCL." *Id.* at 901; *see also id.* at 905 (additionally citing "the heightened risk of related contributors" and the lab's "failure to account for related individuals in close proximity to the firearm" as grounds to preclude testimony). *See also, e.g., People v. Hillary*, Ind. No. 2015-15 (St. Lawrence Cnty. Ct., N.Y., Aug. 26, 2016) (slip op. at 8) (finding an application of STRmix inadmissible under the *Frye* standard because of "the lack of internal validation by the New York State Police crime lab" at the time the laboratory used it in the defendant's case).

Fourth, the sole purpose of calculating a likelihood ratio using STRmix is to assess the relative probabilities that one or more *competing hypotheses* about the contributor(s) to a DNA sample are true. Where the prosecution is the proponent of the evidence, that typically (but by no means always) involves testing a prosecution-supported hypothesis that a defendant and/or alleged victim contributed DNA to the sample, as opposed to some other combination of contributors that tends to negate the prosecution's theory or support the defense's case. But regardless of its particulars, it is widely understood that for a likelihood ratio to "fit" the purpose for which it is offered, an analyst must test competing propositions *that are actually in*

dispute at trial. As the developers of STRmix themselves instructed: "To form an evaluative opinion from a set of observations, it is necessary for the forensic scientist to consider those observations in the light of propositions and forensically relevant case information. The propositions should represent the positions of the different participants in the legal process. In a criminal trial, the propositions will represent the positions of prosecution and defen[s]e, respectively." J. Buckleton et al., *Helping Formulate Propositions in Forensic DNA Analysis*, 54 Sci. & Just. 258, 258 (2014). The government's experts concur, and use STRmix for precisely this purpose. *See* Background Part I.D., *supra* (discussing Green Tx. at 10–11 and Kalafut Tx. at 126–28 on how competing hypotheses are formulated).

B. The "Competing Hypotheses" for STRmix to Calculate

Here, the only competing hypotheses Green was able to input into STRmix involved scenarios in which the DNA profiles she observed were contributed by the government's "persons of interest" (Jane Doe and/or Tyler Johnston) as opposed to unknown, *unrelated* persons. Conditioning each run on the assumed presence of Monica Johnston, Green yielded results indicating that the three-person DNA mixtures (stains 7, 10, 11, 20, and 25) were "at least 1.0 quintillion times more likely if they originated from Monica Johnston, Tyler Johnston, and [Jane Doe] than if they originated from Monica Johnston *and two unknown individuals*." Report, Def. Hrg. Ex. A at 1695A–96A (emphasis supplied); *see also* Green Tx. at 101 (while Green's report simply indicated "unknown," the hypotheses she ran specifically considered only "unknown, unrelated individuals"). As for what Green believed was a four-

person DNA mixture (stain 23), the resulting profile was "at least 1.0 quintillion times more likely if it originated from Monica Johnston, Tyler Johnston, [Jane Doe], and an unknown individual than if it originated from Monica Johnston and three unknown [and unrelated] individuals." *Id.*

The problem with the government's bid to admit this evidence is that Green's statistics have no relationship whatsoever to the parties' actual disputes over the source(s) of DNA in these stains and the resulting inferences they will ask the jury to draw. As defense counsel made clear at oral argument, at this trial, "no one is talking about strangers" as alleged perpetrator(s) of Jane Doe's abuse, nor as potential contributors to the DNA on the comforter. Oral Arg. Tx. at 46. For "the defense hypothesis is not that there was an unknown unrelated person that came in and sexually assaulted Jane Doe on the bed." *Id.* at 72. Instead, "[the defense's] hypothesis has to do with there being unknown *related people*" included in these DNA mixtures of three, four, or five contributors. *Id.* Specifically, the defense contends that there is a substantial probability that Jane Doe's three younger siblings – who, like Doe, had regular innocent contact with the bedspread – were also contributors to the same areas of the comforter that Green tested.[7] And there is no suggestion that

[7] The defense also noted an alternate (but related) hypothesis as it applies to the DNA evidence, even if Green's likelihood ratios were accepted as reliable. According to the defense, the undisputed presence of Monica Johnston in each of these stains makes the results probative of nothing having to do with Jane Doe's alleged sexual assault. This is because, the defense argues, since at least one person who is not accused of any act of sexual abuse is present in these stains (*i.e.,* her DNA comes from one or more innocent contacts she had with the bedspread on her own bed), the DNA makes it no more likely that Johnston or Doe's DNA was deposited during an

any of the Johnstons' other children were victims of sexual abuse. Thus, had Green utilized STRmix to test this hypothesis, the resulting likelihood ratios might have undermined, rather than supported, the government's claim that the DNA results "make it more likely than not," *id.* at 12, that Doe's claim she was sexually abused on the bedspread is true. But as they are, the defense argued, "the likelihood ratios calculated here have no bearing on this case because they don't tell us anything about the actual case situation presented." *Id.*

The Court agrees. The defense's position is grounded in the well-established application of Rule 702 requiring the proponent of scientific evidence to demonstrate an articulable "fit" between that testimony and a disputed issue at trial. *See Daubert,* 509 U.S. at 591. And here, the defense's argument is by no means a far-fetched theory as to how Green might conceivably have approached the task at hand (had USACIL given her the tools to do so). Indeed, a special report by a group of the nation's leading DNA experts recently issued clear guidance to analysts in cases like these, which effectively mirrors the steps that Johnston claims USACIL should have taken here. *See* Expert Working Group on Human Factors in Forensic DNA Interpretation, *Forensic DNA Interpretation and Human Factors: Improving Practice Through Systems Approach*, 65–66 (2024) ("[W]hen DNA is recovered from bedding or clothing

---

act of sexual assault than it does Monica's and "does not support the government's theory." *See* Def. Mem. at 25. However, this argument does not rest on any claim that Green's likelihood ratio calculations were not reliably performed; nor does it challenge the methodology she used to assess the number of contributors to the DNA mixtures. As such, it is more properly viewed as a Rule 403 issue, and it will be addressed in that section of the Court's ruling.

in a familial sexual assault case, relevant questions include: How many first-order relatives (both adults and children) live or have regular access to the home? . . . When elimination samples are not available, the analyst should seek additional case information on who could be an alternate source of the DNA and carefully consider the resulting propositions. If meaningful to the case, *the alternative should involve a relative and not just an unknown, unrelated individual*.") (emphasis supplied) ("*Human Factors*").

Here, Green's "1 in 1.0 quintillion" likelihood ratios considered only the government's "persons of interest" against potential combinations of Monica Johnston and two or three irrelevant *unrelated* persons. As such, her statistics will not aid the jury in resolving the parties' competing theories as to which three or four (or more) persons' DNA is in these areas of the comforter. Accordingly, the government has failed to meet its burden under Rule 702. *See Daubert*, 509 U.S. at 591 (to be admissible, the proponent of scientific evidence must show that it will "assist the trier of fact to understand or determine a fact in issue" (quoting Fed. R. Evid. 702)).

C. USACIL Could Have, But Did Not, Internally Validate STRmix to Calculate Likelihood Ratios Where the Potential "Unknown" Contributors Are Biological Relatives of the Person(s) of Interest

This is not to say that STRmix can never be used to calculate likelihood ratios where one party seeks to determine the probability that unknown, related persons are potential DNA contributors, including but not limited to alleged intrafamilial crimes. The government's expert, Tim Kalafut, has written about scenarios in which

probabilistic genotyping can and should be used to calculate likelihood ratios that include such propositions: for example, where one party has hypothesized that "a person of interest's brother [who is] unavailable for DNA analysis is the source of the DNA." Kalafut Tx. at 126, 128.

Indeed, the developers of STRmix have incorporated a feature into the software that permits analysts to use it in precisely this way. *Id.* at 128. But unlike many other forensic DNA laboratories,[8] USACIL has not internally validated this function of STRmix. As a result, as Kalafut explained, USACIL "doesn't have that [unknown related person] function of STRmix turned on," and Green was unable to use it here. Kalafut Tx. at 129. *See also* Background Part I.D., *supra*; Sutton Aff. ¶ 24 ("[a]s it relates to biological relatives, USACIL does not currently evaluate DNA results to distinguish biological relatives using STRmix. Although these calculations can be done using STRmix, USACIL has not internally validated STRmix for this purpose.")

[8] For example, the Las Vegas Metropolitan Police Department ("LVMPD") allows for likelihood ratio statements that "[c]ompare[] the probability of observing the evidence given the person of interest is a contributor to the DNA profile vs. the evidence originating from an *untested defined relative of interest (e.g., brother, grandparent, niece, uncle, etc.).*" Las Vegas Metropolitan Police Department Forensic Laboratory, Biology/DNA Detail Procedures Manual, at 253 (2024), ECF No. 103-20. However, even the LVMPD's own internal validation study, which included samples of related individuals, warns that "[m]ixtures comprising DNA of related individuals and comparison of such mixtures to other related individuals is a known limitation of any DNA interpretation system. In this scenario, a close relative of the true contributor could be included purely by the spread of genotype combinations and the fact that they are likely to possess some of the same alleges as the true donor(s)." Las Vegas Metropolitan Police Department, Internal Validation of STRmix v2.6 (QIAGEN Investigator 24plex QS with 3500xl), at 7 (2020), ECF No. 103-17.

The government does not dispute that USACIL lacked the necessary validation to include the defense's hypotheses in any of Green's likelihood ratios. But it argues that "USACIL's lack of validation testing for biological relatives has no effect here" because Tyler Johnston and Jane Doe "are not biological relatives of any kind" and "the evidence here is not comprised of complex mixtures of first-order biological relatives." Gov. Sur-Reply at 2, ECF No. 144.

This argument wholly misses the mark. It is, of course, the government's *contention* that the stains Green tested do not consist of "complex mixtures of first-order biological relatives," and that two people who are "not biological relatives" (Johnston and Jane Doe) are present in each of these stains. But the purpose of calculating a likelihood ratio is to help the jury assess whether the government is correct: that is, to use science to measure the probability of the government's contention being true *against a competing hypothesis*. Here, the defense seeks to use STRmix's "unknown, related persons" feature to do precisely that. It would have STRmix calculate likelihood ratios that consider the relative likelihood that Doe's three half-siblings are contributors to these stains – either instead of, or in addition to, Doe and/or Johnston. All three of these "unknown, related" persons are Doe's "biological relatives." All four children are "first-order biological relatives" of the one person – their mother – whose DNA Green affirmatively "matche[d]" to the stains on the bedspread. And two of the four children are "first order biological relatives" who share all their DNA alleles with Monica and Tyler Johnston.

In sum, it is widely accepted in the forensic DNA community, including by each of the government's experts, that when an DNA analyst uses probabilistic genotyping to calculate a likelihood ratio for a potential inclusion, the defense is entitled to have its reasonable competing hypotheses included as alternate propositions. USACIL did not do so here, because it lacked the necessary validation to use the "unknown, related person" function of STRmix. That Green could not perform the analysis required for the parameters of this particular case was no fault of hers. But that does not exempt the government from its burden of proving that her analysis would be "helpful to determine a [disputed] fact in issue," and "reflects a reliable application of [DNA] principles and methods to the facts of this case." Fed. R. Evid. 702(a), (d). The government has not met either burden here.

II. With the Limited Information Provided to Her, Green Was Unable to Make a Scientifically Reliable Determination of the Number of DNA Contributors to a Common Bedspread in a Six-Member Household That Included Numerous First-Degree Relatives

Johnston argues, as a separate and independent ground for exclusion, that the government has not shown that Green's assessment of the number of contributors ("NOC") to each stain is of "sufficient scientific reliability." Oral Arg. Tx. at 38. The Court agrees. The NOC determination is a "foundational part of every calculation" that probabilistic genotyping software performs, and "[i]f that input is in doubt, the reliability of the entire analysis is necessarily in doubt." *United States v. Williams*, 382 F. Supp. 3d 928, 937 (N.D. Cal. 2019). Because Green's NOC determination was both "integral to the mixture interpretation process," Def. Reply Mem. at 13, and to

the inferences the government will ask the jury to draw from the DNA testimony, the Court must preclude Green's testimony on this ground.

The Court reaches this conclusion for several reasons, including: (1) USACIL has not conducted internal validations to measure the error rates of its analysts in casework reflecting the kinds of challenging conditions seen here – in particular, the difficult task of distinguishing among DNA donors to a common household item when numerous persons in that household came into repeated (or frequent) contact with the item, and when that pool of potential donors will exhibit high allele sharing due to the presence of numerous first- and second-order relatives; (2) because reference samples from three of the six Johnston family members were never provided to Green, she was unable to follow even the limited protocols USACIL did have in place for reliably distinguishing among related persons; (3) the government's own consulting expert concluded that Green overlooked at least one "masked," additional DNA donor in a mixture that likely comprises five (not four) individuals, further undermining confidence in the reliability of her other NOC determinations; and (4) counsel for the government did not conduct basic inquiries to determine the nature of these untested relatives' contact with the bedspread it submitted for testing in 2021 until after the *Daubert* hearing had concluded in March 2025, depriving Green of key information about the children's contact with the bed that may well have changed her core assumptions about "expected" donors to the mixtures and impacted her subsequent analyses.

A. Challenges in NOC Determinations in Mixed DNA Samples with High Allele Sharing Among Multiple Potential Donors

It is well understood that "DNA analysis of complex mixtures is inherently difficult." PCAST Report at 8. *See also* Murphy, *Inside the Cell* at 91 ("[A]nalysts may think they are dealing with a straightforward two-person mixture, only to learn that in fact a third contributor's profile is hidden beneath.").

> Such samples result in a DNA profile that superimposes multiple individual DNA profiles. Interpreting a mixed profile is different from and more challenging than interpreting a simple profile, for many reasons. It is often impossible to tell with certainty which genetic variants are present in the mixture or how many separate individuals contributed to the mixture, let alone accurately to infer the DNA profile of each one.

PCAST Report at 8.

The already-challenging task of determining how many persons contributed to a DNA mixture becomes even more difficult when there are related individuals contributing, or potentially contributing, to the evidence. For one of the biggest challenges in distinguishing among individuals and identifying the NOCs is "allele sharing." *See* Butler, *Interpretation* at 153 (citing, as a "primary complicating factor[] in deciphering mixture components[,]" the "potential for allele sharing that results in peak signal stacking" and which may mask the presence of additional donors to the mixture). While unrelated persons are likely to share at least some DNA alleles by chance, allele sharing is particularly common among biological relatives, particularly (but not only) among first-order relatives such as parent-child and full-sibling relationships. *See* Kalafut Tx. at 94; Green Tx. at 23–24.

That even experienced DNA analysts who are asked to determine NOCs in complex mixtures too often get it wrong is well understood. Controlled studies have shown that even when using standard protocols, the risk that a DNA analyst will underestimate the true number of contributors to a DNA mixture is not insignificant. And the underestimation of NOCs may also lead the analyst to falsely include an innocent person or falsely exclude a true donor. For example, in 2015, the national "MIX13" study, which secured the participation of DNA analysts from 108 laboratories across 48 states, asked the subjects to assess the number of contributors and compare the mixture against multiple reference samples in cases where the "ground truth" was known. One of the casework samples was a four-person mixture that researchers knew contained high allele sharing among donors, purposely increasing the risk that the analysts might mistakenly deem it a two-person mixture. Fully 69% of participants who were given this four-person mixture made a false positive error – meaning that they falsely included an "innocent" subject who was not an actual donor to the mixture. *See, e.g.,* M. Coble et al., *Interpretation Errors Detected in a NIST Interlaboratory Study on DNA Mixture Interpretation in the U.S.* (2015), available at https://www.nist.gov/document-8; G. Dembinski et. al., *Estimating the Number of Contributors of Theoretical Mixture Profiles Base on Allele Counting: Does Increasing the Number of Loci Increase the Success Rates of Estimates?*, 33 Forensic Sci. Int'l: Genetics 24, 26–29 (2018) (subjects using a maximum allele count method underestimated NOCs in known five-person mixtures at a rate of 43%, and in six-person mixtures at a rate of 92%); M. Coble et. al.,

Uncertainty in the Number of Contributors in the Proposed New CODIS Set, 19 Forensic Sci. Int'l: Genetics 207, 209 (2015) (study finding that at least 86% of known six-person mixtures were misidentified by analysts as having five or fewer contributors).[9]

An analyst's erroneous assessment of the NOC can have significant consequences, including "rippling effects through subsequent stages of analysis." K. Kwong, *The Algorithm Says You Did It: The Use of Black Box Algorithms to Analyze Complex DNA Evidence*, 31 Harv. J. L. & Tech. 275, 291 (2017); *see also* Murphy, *Inside the Cell* at 95 ("[t]he more complex the mixture, the more complicated the array of possible explanations that do *not* include inculpating [the POI]"). That is particularly so in the era of probabilistic genotyping, since an analyst must input a specific NOC (rather than a minimum or "at least X . . ." estimate) into STRmix and

[9] In its original motion, the defense also cited the results of Kalafut's own 2022 study, which reported an overall error rate of 25 to 33 percent among analysts who were asked to determine the NOCs in known three-person mixtures where the samples included a high degree of kinship. Def. Mem. at 35 and Def. Mot. Ex. G, ECF 103-7; Kalafut Tx. at 88–89. Having heard extensive direct and cross-examination of Prof. Kalafut on this issue, the Court is not convinced that the differences between the bedspread stains in this case and the known "first-order relatives" mixtures in Kalafut's study are as different as Prof. Kalafut himself indicated when he began his testimony (describing them as "apples and oranges," *id.* at 23). That is particularly so in light of those aspects of the defense's cross-examination that explored the data in greater detail (such as when Kalafut agreed Green found "balance" between two of three donors in certain stains, and which more closely parallel his own research). Nevertheless, the Court has no reason to doubt that Kalafut has a more fulsome understanding of the data and scope of his own research than does this Court. For that reason, it does not rely on Kalafut's study in this opinion, other than for the general (and undisputed) proposition that in at least some DNA mixtures that involved high allele sharing among biological relatives, his study found a significant risk of error in the subject analysts' NOC determinations.

similar programs for the software to generate a likelihood ratio, and the program must take as a given the analyst's predetermined NOC when it calculates the likelihood ratio. *See* Jo-Anne Bright, et al., *Internal Validation of STRmix—A Multi Laboratory Response to PCAST*, 34 Forensic Sci. Int'l: Genetics 11, 22 (2018) (noting that "overestimation of the number of contributors generally leads to lower *LR*s for true contributors and an increase in *LR*s for non-contributors" and that "underestimating the number of contributors can result in false exclusions of true donors").

The potential impact of an unreliable NOC determination in this case is not hard to discern. First, Green's likelihood ratios depend on her opinion that five of the six stains contained only three contributors. Second, from this purportedly limited universe of donors, the government will ask the jury to infer that DNA from Doe was deposited in these specific areas of the bed at the same time as was Johnston's (and that his DNA came specifically from seminal fluid) – making it, in the government's view, more likely that Doe's account of being raped in this area of the bed several days earlier is true than not true. Oral Arg. Tx. at 12. (And while not terribly specific on how it intends to reconcile that theory with the fact that Monica is also a contributor, it appears that the government will argue that her DNA was separately deposited on an unrelated occasion, such as marital intercourse with Johnston or sleeping in the bed.) But if Green's NOC determination is wrong, and the "true number" of NOCs is equally or more consistent with the defense hypothesis that some amount of DNA from *all of the Johnstons' children* – not just Doe – is on the bed,

including in these five stains, it would substantially weaken the scientific foundation on which the government rests its proffered inferences.

Of course, the fact that NOC determinations are inherently subjective, and even experienced DNA analysts may disagree on the "correct" NOC for a complex mixture, is not itself a basis for exclusion. Courts have routinely allowed analysts to testify to their STRmix-generated results, notwithstanding defense attacks on the overall subjectivity of the NOC determination that the analyst used to obtain the likelihood ratios. *See, e.g., Tucker*, 2020 WL 93951 at *4 (noting that courts have "overwhelmingly admitted expert testimony based on STRmix results" and doing same in case involving "single source and simple mixtures," while "acknowledg[ing] the concerns surrounding the fast-expanding use of probabilistic genotyping"). In a recent unpublished decision, for example, the Second Circuit found that the district court did not abuse its discretion in denying a *Daubert* hearing and admitting STRmix evidence at trial, where defendant had characterized the laboratory's method of calculating NOCs as one based solely on "a technician's whims," yet the defendant "presented no evidence suggesting that the government's method lacked accuracy and dependability." *United States v. Green,* 2024 WL 4488577, at *2 (2d Cir. Oct. 15, 2024) (summary order). But for the reasons outlined above, this case is different.

B. USACIL Lacked the Internal Validation Needed to Establish that Green Reliably Determined the Number of Contributors on a Shared Item That May Have Contained DNA from Six Related Persons; and Green Was Not Given the Information Needed to Follow the Alternate Protocols USACIL Did Have

The Supreme Court has instructed that where a scientific expert's use of a particular technique is challenged, the proponent of the evidence will typically need to provide the court with information about the methodology's "rate of error:"

> [I]n the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error, see, *e.g., United States v. Smith,* 869 F.2d 348, 353–354 (CA7 1989) (surveying studies of the error rate of spectrographic voice identification technique), and the existence and maintenance of standards controlling the technique's operation, see *United States v. Williams,* 583 F.2d 1194, 1198 (CA2 1978) (noting professional organization's standard governing spectrographic analysis), cert. denied, 439 U.S. 1117 (1979).

Daubert, 509 U.S. at 594.

In the field of forensic DNA analysis, measurement of known or actual error rates that accompany a laboratory's methodology is done through internal validation. It involves testing under controlled conditions that are representative of the kinds of actual cases that may be the subject of an analyst's in-court testimony. *See, e.g., Human Factors* at 150 ("Since the party proffering the evidence bears the burden of demonstrating reliability or general acceptance, that party and the expert *should be prepared to demonstrate that internal validation involved ground-truth-known samples that are representative of the types of samples the analyst interprets in their casework and, preferably, also in the case at hand*." (emphasis supplied) (citation omitted)). Crucially, "an internal validation should aim to test the gray areas of a measurement *to ensure challenging, complex, or unusual samples that may arise in*

casework are covered within the FSSP's internal validation." *Id.* at 201 (emphasis supplied). For this reason, "[w]hen considering the admissibility of testimony about complex mixtures (or complex samples), judges should ascertain whether the published validation studies adequately address the nature of the sample being analyzed." President's Council of Advisors on Science and Technology ("PCAST"), *An Addendum to the PCAST Report on Forensic Science in Criminal Courts*, at 9 (2017), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf).

However, USACIL has conducted no internal validations to ensure that analysts like Green are making reliable determinations in casework that poses the kinds of challenges presented here. As USACIL's technical leader acknowledged, "[a]s it relates to biological relatives, USACIL does not currently evaluate DNA results to distinguish biological relatives using STRmix. Although these calculations can be done using STRmix, USACIL has not internally validated STRmix for this purpose." Sutton Aff. ¶ 24. More fundamentally, it has not otherwise validated protocols for handling DNA mixtures that exhibit high allele sharing due to the contributions of various biological relatives. Other labs have designed their validation studies to assess the impact of relatedness on DNA interpretation. For example, the Jefferson County, Colorado Regional Crime Laboratory (one of the few laboratories to make its internal validations available to the public) included known DNA mixtures of related persons in its validation studies, measured the analysts' proficiency in that subset of cases, and used the results to adopt protocols that *may*

more reliably account for those confounding factors when calculating NOCs and likelihood ratios. *See* Jefferson County Regional Crime Laboratory, Internal Validation of STRmix V2.6 for the analysis of GlobalFiler profiles, 21–31 (2019), ECF No. 103-16.

Because USACIL has not done so, its procedures address the challenges of complex related-person mixtures only in the most general terms. *See* DFSC DNA 114.1 § 6.7 (warning USACIL analysts that "mixtures involving biological relatives that cannot be deduced into individual contributors *should be interpreted and reported with caution* due to allele sharing") (emphasis supplied). Nor is there any specific guidance in the USACIL protocols that Green cited in her testimony as applicable to this case, which merely instruct analysts to "[l]ook for any peak height imbalance at multiple loci for likely contributors which would indicate an additional contributor may be present." DFSC DNA 114.1 § 6.3.2. This directive applies to the assessment of NOCs in *all* mixture interpretations. It does not address the additional, known challenges of interpreting mixtures where high allele sharing is almost certain to be present — such as items collected from households with many first-order relatives, in mixtures that already have clear indications of three or more contributors.[10] And it certainly provides no indication that USACIL has measured its analysts' proficiency and error rates when analyzing mixtures of this kind.

[10] As Kalafut affirmed, simply looking for peak height imbalances and applying the 50% ratio "rule of thumb" is not a sufficiently reliable method of determining NOCs in this category of cases. For when "analyzing mixtures with a high degree of allele sharing, like those involving relatives, additional contributors if they're

In the government's affidavit from USACIL's Technical Leader Sutton, he noted that even though USACIL lacks any specific validations of these methods, "USACIL uses other mechanisms" to distinguish among biological relatives who may be additional DNA donors. These specifically include "requesting references from all these individuals to compare their DNA to the biological evidence in question." Sutton Aff. ¶ 24. However, Green was unable to utilize even that mechanism here. She was never given DNA standards from fully half the members of the Johnston household (each of whom are related to Monica, Tyler, or Doe herself), even though Green's case manager specifically requested them from the CID at the outset. For that reason, as Green herself explained, her work was "limited by the references [she] received." Green Tx. at 157.

Recognizing that USACIL lacked any specific protocols or proven methodologies for testing the defense hypothesis here – that untested children of the Johnstons' might well have been the fourth, fifth, or sixth contributors to these stains – the government urged the court to find that, as a practical matter, USACIL's lack of applicable validations and protocols is of no consequence. On this point, the government relies heavily on Kalafut's agreement that there were, in the government's words, "no unaccounted-for alleles" in Green's data (other than in stain 23, *i.e.,* the four- or five- person mixture). Oral Arg. TX. at 69. It appears that the government's view is that (other than in Stain 23), since Green detected no alleles

---

biological relatives can still be masked even if the peak height ratios are within this [50%] rule of thumb." Kalafut Tx. at 92.

that *must* have come from someone other than the three known individuals, there was little chance that Green had actually failed to detect additional donor(s).

Yet Kalafut quickly qualified this statement by adding that there was "a little more to it" than simply identifying alleles and finding them to be present in the known donors' profiles. Kalafut Tx. at 155. (That appears to be especially true here, where two of the three untested children share *all* of their parents' alleles, so any DNA "accounted for" in the parents' profiles would apply equally to their children.) He explained that an analyst's assessment of peak height ratios and whether they indicate a potential masked donor also factored heavily in this assessment. *Id.* at 154–55. And he acknowledged that whether an observed peak height ratio of 50% or greater indicates the presence of additional minor donors is based on a general "rule of thumb" that and is not a fixed standard among DNA laboratories and "incorporates uncertainty." Kalafut Tx. at 52, 91.

The government's insistence that Green's data reliably forecloses the possibility of additional donors, to the point where the court should overlook the fact that it never validated her methods internally, is also difficult to square with the recent history of this case. Before the *Daubert* hearing, the government expressed not just confidence, but certainty, that Green had identified the maximum number of contributors to each of the stains at issue. *See* Gov. Opp. at 66 ("[*I*]t *is not possible* that any such hypothetical DNA could be fully masked here. Moreover, the actual data in this case *makes clear* that there is no additional DNA from these children in the mixtures tested.") (emphasis supplied). But when Kalafut – a consulting expert

retained by the government to help defeat the defense's *Daubert* challenge – was called to testify, he informed the Court that in his view, the presence of additional DNA from "a biological child" of the Johnstons' in at least one of the mixture stains tested by Green was not just possible, but was the "most likely" explanation for the data. Kalafut Tx. at 38.

Of course, the fact that the government's experts disagree on one or more conclusions from the same data does not render Green's testimony inadmissible. *See In re Fisher-Price Rock 'N Play Sleeper Mktg., Sales Pracs. & Prods. Liab. Litig.,* 567 F. Supp. 3d 406, 412 (W.D.N.Y. 2021). But combined with the lack of other protocols and methods that would provide more objective assurances that Green's NOC conclusions in this case were reliable, the fact that Kalafut identified at least one instance in which Green had failed to detect the contributions of a masked biological child – weeks after the government assured the Court that such an error was "not possible" – gives further reason to find that the government's burden has not been met.

Finally, again through no fault of Green's own, the record now contains additional information about expected contributors to the comforter that Green did not have the opportunity to consider before she analyzed the data, determined the NOCs, and issued her report. This post-hearing Rule 3500 disclosure included a statement from Jane Doe that her two youngest siblings co-slept in their parents' bed. 3500-JD1-050. In fact, the government's handwritten notes from its recent interview with Doe initially reported this co-sleeping as a "sometimes" occurrence but then

corrected it to specify that it happened "every day". *Id.* Yet at the time Green conducted her analysis, she was aware only of Monica Johnston's statement "that the children would be on the bed at some points but [it] would be 'harmless fun such as the children wrestling on it and playing around.'" Case Management Notes, Def. Hrg. Ex. A at 1836A. That was important information (and led the case agent to immediately request the children's reference samples). But in Green's mind, it did not rise to the level at which she would expect these children's DNA to be found on the comforter, at least not to the extent she did for the two adults who regularly slept in that bed. *Compare* Green Tx. at 172 (she knew it was Monica's bed, so "there was an expectation her DNA is already on this comforter") *with* Green Tx. at 154–55 (Green had no information about who, if anyone, co-slept with Monica and Tyler).

It was Green's "reasonable expectation" about Monica's DNA being present on the comforter that informed her decision to "condition," or presume, Monica's contributions when interpreting the remaining DNA in the mixtures. *Id.* 173. And both she and Kalafut had a similar "expectation" as to Johnston's likely contributions, *i.e.,* as the other adult who slept in the bed nightly, they presumed his DNA would be on the comforter stains, and thus it was reasonable for Green to condition several hypotheses on that assumption when she ran various hypotheses in STRmix. *See id.* at 63 ("It's [Monica's] own bed. And so assessing her presence in those mixtures was the next logical step"); Kalafut Tx. at 98 ("I would expect to find Mr. Johnston and Mrs. Johnston on that comforter. I expect to find me on my own comforter").

In sum, the expectation that both Monica and Tyler were likely donors to each of the bedspread stains Green tested informed her analysis at all steps of the process. Green herself recognized the potential implications of any "conditioning" assumption when she asked Kalafut via email whether she should be "concerned for court" because she had conditioned all of her STRmix hypotheses on the presence of Monica's DNA. That may have been an entirely reasonable assumption for her to make, given that the evidence was collected from the bed where Monica slept. But what Green did not know was that there were two *additional* family members who were also sleeping in that same bed every single night – each of whom shared 100% of their DNA alleles with Tyler and Monica, and with whom Jane Doe shared at least 25% of her own alleles.

This belated, post-hearing disclosure of highly relevant factual background deprived the defense of the chance to inquire about how, if at all, this information might have changed Green's analytic process had she been made aware of it from the outset. For example, Green was never asked whether, if she knew that the youngest Johnston children slept in the bed with their parents nightly, she would have similarly "expected" their presence in the mixtures; how, if at all, she would have accounted for that expectation in her NOC assessment or use of STRmix; or even whether she would have been comfortable reporting her results at all if she lacked DNA standards from half the people who slept in the bed around the time of its collection. The Court will not speculate about Green's answers to questions that she never had an opportunity to be asked. But this history provides yet another reason

to be concerned that Green lacked "sufficient facts or data" to guide her testing and analysis.

* * *

In sum, this presents the rare case in which uncertainties about the reliability of a DNA analyst's conclusions are so numerous and consequential that the Court would fail in its gatekeeping function if it did not exclude the analyst's testimony from trial. For the reasons stated above, the government has failed to establish that Green's determination of the number of contributors in the stains she tested was "based on sufficient facts or data," Fed. R. Evid. 702(b), or that it "is the product of reliable principles and methods . . . [as applied to] the facts of the case," Fed. R. Evid. 702 (c), (d).

III. Even if Scientifically Reliable, the DNA Evidence Would Nonetheless Be Inadmissible Under Rules 401, 403, and 702(a)

The Court also finds that, even if Green's DNA results were sufficiently reliable to be admissible under *Daubert*, Johnston has made a compelling argument for preclusion under Fed. R. Evid. 401, 403, and Rule 702(a).

The complete, post-hearing record makes clear that before USACIL was provided with the Johnstons' marital bedspread, Jane Doe, Tyler Johnston, and Monica Johnston each had contact with that item in numerous ways that had nothing whatsoever to do with any alleged sexual assault. And there is remarkable unanimity among the parties' three experts that (1) Green's testing provides no information that can help answer the question of how, when, and under what circumstances any DNA from these three people was deposited on the bedspread, and (2) even if semen was

present in the tested areas, Green's testing cannot identify what substance(s) – semen, saliva, skin cells, etc. – actually produced the DNA profiles she detected. Thus, as a scientific matter, Green's findings make the defense's explanation for the presence of Doe's and Johnston's DNA in these areas (that they were separately deposited during each person's ordinary contact with the bed) and the government's hypothesis (that they were deposited when Johnston raped Doe on the bed) *equally likely to be true*. As such, Green's testimony simply has no "tendency to make a [disputed] fact *more or less probable* than it would be without the evidence," Fed. R. Evid. 401(a) (emphasis supplied) – at least, not any "fact [that] is of consequence in the action," Fed. R. Evid. 401(b).

Further, even if the Court were to find that it met the low bar of relevancy under Rule 401, any probative value Green's testimony would add to the government's case is plainly outweighed by the significant danger that the DNA evidence would confuse the issues and mislead the jury. *See* Fed. R. Evid. 403; *see also* Fed. R. Evid. 702(a) (expert testimony is admissible only if the proponent demonstrates that it is "more likely than not" that it "will help a trier of fact to understand the evidence or determine a fact in issue").

The legal standards that govern this question are not in dispute. "To be relevant, evidence need not be sufficient by itself to prove a fact in issue, much less to prove it beyond a reasonable doubt." *United States v. Abu-Jihaad*, 630 F.3d 102, 132 (2d Cir. 2010). Accordingly, relevance under Rules 401 and 403 has been described as a "low threshold, easily satisfied." *United States v. Gramins*, 939 F.3d

429, 450 (2d Cir. 2019). Additionally, just because proffered evidence "might be construed . . . innocently" does not mean that the evidence is not relevant. *Abu-Jihaad*, 630 F.3d at 132. Like lay testimony, scientific testimony or data "need not be conclusive in order to be relevant." *Contemporary Mission v. Famous Music Corp.,* 557 F.2d 918, 927 (2d Cir.1977). However, evidence that is not conclusive must still make "a proposition more probable than not" to be admissible. *United States v. Schultz*, 333 F.3d 393, 416 (2d Cir. 2003) (quoting *S.E.C. v. Singer*, 786 F. Supp. 1158, 1166 (S.D.N.Y. 1992)). Thus, the proponent of the evidence must still show that, if credited by the jury, it would create some increased likelihood, however slight, that the asserted proposition is true.

A. Rule 401 and 702(a)

Despite the technical complexity of the government's DNA evidence, Johnston's challenge under Rules 401 and 702(a) is a simple one. Even assuming the jury credits all of Green's testimony and affords the government the benefit of all reasonable inferences from it, the defense argues, "[w]hat the scientific evidence says is Jane Doe was on the bed, Mr. and Mrs. Johnston were on the bed, at some point semen got on the bed. That's what we have. And you can't draw more from it than that." Oral Arg. Tx. at 34. The defense argues that these facts, even if true, "do[] not get us anywhere near sexual assault" – in other words, they make it no more likely that Johnston sexually assaulted Doe on that bed than that they each had innocent contact with the bed on any number of other occasions. *Id.* at 33. Further, the defense argues, this is so not just because Green affirmed that her testing does not answer

such "activity-level propositions" (*i.e.,* the how, why, and when of DNA deposits). It is also because Green found DNA consistent with Monica Johnston, in each of these stains, as well as (at least one) other biological child of the Johnstons', none of whom had anything to do with any alleged sexual assault. *Id.* at 33–34.

The government, by contrast, argues that "it would be an appropriate inference from the testimony of Forensic Biologist Green that Doe is, in fact, a contributor to a certain number of mixtures of DNA in which the Defendant's semen is present that were recovered from the very places Doe alleges the Defendant raped her, *making it more likely than not the Defendant raped her in those places*." *Id.* at 12 (emphasis supplied). The Court asked the government to explain more precisely on what basis it had to assert – and ask a jury to find – that it is "more likely" that these results are due to an act of sexual assault than to any other combination of DNA-transfer circumstances. *Id.* at 13–14. The government responded that "[i]f a child's DNA is mixed with a defendant's semen, I think it's a fair basis to say it is more likely than not – there's a reason they mixed. And the Government submits that reason is he was sexually assaulting her in that location." *Id.* at 15.

The Court then pressed the government as to whether there was any evidence from the *Daubert* hearing or the extensive forensic literature cited by both parties that supports the government's "more likely than not" inference. The government pointed to Green's testimony that "semen is a rich source of DNA," and that "Doe was in certain of these stains a major contributor." *Id.* at 18. According to the government, "[t]hat means that Doe is contributing even more DNA than the rich

supply of semen deposited by the Defendant." *Id.* The composition of these stains, according to the government, corroborates what it anticipates Doe will testify to at trial about her contact with the bedspread: namely, that she "did not frequently spend time in this bed in . . . an innocent manner,"[11] but instead "that she was primarily in this bed to be abused by the Defendant." *Id.*

There are two fundamental problems with the government's claim. The first is that it is counterfactual (or at least, incomplete and misleading). It is true that Green answered "yes" when asked by the government if "semen contain[s] a rich quantity of DNA." Green Tx. at 20. But Green was not asked whether seminal fluid yields any higher (or lower) concentration of human DNA than any other biological material that may also have accounted for the presence of these individuals' DNA on the bed, such as saliva, sweat, skin cells, vaginal fluids, or other bodily material. In addition, of the five stains that Green identified as three-person mixtures, she concluded that Doe was the "major donor" in "at least one, maybe two or three" of them. *See* Green Tr. at 187. And in some of those, only a trace amount of Johnston's DNA was detected (at levels that Green pegged to be as low as 2% of the total contribution in one sample). *See* Gov. Hrg. Ex. 252 at 8. Thus, if the government were correct in its view about the impact of the "rich" quantity of DNA in semen, the stark contrast between Doe's and Johnston's relative contributions would seem to

[11] The Court notes that this was counsel's best recollection of Doe's earlier statements (if any) regarding the frequency and nature of her other contact with the bed, but pre-dated the government's March 11, 2025 interview with Doe in which she specifically reported spending time on the bed with her siblings to use her personal electronic device.

suggest an entirely different time and manner of deposit (*i.e.,* they would not both have come from an allegedly recent act of sexual assault by Johnston against Doe). In addition, as defense counsel noted at oral argument, in some of the mixtures Green tested, Mr. and Mrs. Johnston are "quite proportional in their contributions." Oral Arg. Tx. at 72–73. Were the government's proposed lens an appropriate way to assess the time and manner of deposit, then, the data would appear to be more consistent with the Johnstons' having marital intercourse on the bed together, and Doe depositing her own (higher or lower) amount of DNA at a separate, unrelated time.[12]

The second – and much more significant – problem with the government's argument is that it cannot be reconciled with the testimony of its own experts. Indeed, it is directly contrary to their testimony on this very issue. *See, e.g.,* Green Tx. at 138 (agreeing that nothing in her results allowed her to draw any conclusion as to "whether Jane Doe's [DNA] was deposited at the same time or a different time" as Tyler's and Monica's); Kalafut Tx. at 132 (affirming that Green's testing did not in any way "address[] the question of how the DNA may have been deposited" on the bedspread).

Green also affirmed that the presence of semen indicates only that there were some DNA contribution(s) from a male at some point in time. *See* Green Tx. at 138.

[12] Because there is now undisputed evidence on record that Doe had direct contact with the comforter on various occasions unrelated to any alleged sexual assault prior to its collection as evidence, the Court need not consider Johnston's additional argument that, given the extraordinary sensitivity of modern DNA testing, "indirect transfer" of Doe's DNA – for example, through contact with her siblings or parents – might also have accounted for the presence of Doe's DNA. Def. Mem. at 24; *see also* Burrill Aff., Def. Reply Mem., Ex. FF ¶¶ III.2, III.8, ECF No. 135-1.

But Green explained that she cannot identify "the actual activity of how [the DNA] got on the comforter." *Id.* Most importantly, her data indicates nothing about whether any DNA alleles consistent with Monica, Doe, and even Tyler Johnston himself came from contact with the comforter at the same time as any semen in those stains, *or at different times*. *Id.* In this respect, Green's testimony was fully in line with not just Kalafut's views, but also the opinion of the defense's expert, Dr. Julia Burrill, who also reviewed Green's report, data, and case file:

> The timing or extent of any previous contact by parents or by any of the four cohabitating children is unknown. The persistence of any previously deposited DNA on the comforter is also unknown, meaning it cannot reliably be inferred how any particular DNA within a mixture got there or how long it had been there, or if distinct profiles of DNA were deposited concurrently. . . . [Thus], there is no basis for a conclusion that a specific contact, action, or behavior resulted in the DNA recovered and tested in this case.

Burrill Aff., Def. Reply Mem., Ex. FF ¶ III.3 (emphasis supplied), ECF No. 135-1.

For these reasons, the government's claim that the DNA results show the "more likely" means in which Doe's and Johnston's DNA was deposited on the comforter is pure speculation. It has no basis in science, nor in the testimony of the government's own experts.

The Court does not suggest that the government intentionally misrepresented its experts' findings or the inferences they support. But clearly, the government failed to understand – or perhaps had great difficulty acknowledging – the limitations of Green's DNA testing and data, even after all three experts explained as much. Regardless, the record is clear that Green's DNA results simply have no "tendency" to make a disputed fact – namely, whether Doe was sexually assaulted on her parents'

bed – any "more or less probable than it would be without the evidence." Fed. R. Evid. 401.

B. Rule 403 and Rule 702(a)

Even if the Court were to conclude that the government's proffered evidence has "the required indicia of scientific reliability," *Nimely*, 414 F.3d at 396, and provides some evidence upon which a reasonable jury could rely to resolve the factual dispute over how Doe and Johnston's DNA may have come to be present in these stains, the Court must still find that its probative value is not "substantially outweighed by . . . unfair prejudice, confusing the issues, [or] misleading the jury[.]" Fed. R. Evid. 403. Here, even assuming Green's testimony made the government's contention that Doe was assaulted on the comforter even slightly more likely to be true than it would be in the absence of that evidence, such probative value would be substantially outweighed by the risk that it would both mislead and confuse the jury.

This is so for several reasons. First, as noted above, even learned counsel for the government do not appear to understand key limitations of the testing Green performed, *i.e.,* it only speaks to *whose* DNA might be on the comforter but says nothing about how or when it got there. If the government remains unable or unwilling to accept those limitations even after spending months immersed in the data and conferring with its experts, the Court cannot have confidence that a lay jury would be able to stay within those bounds. The Court agrees with defense counsel that there is a significant risk that the jury will simply "put together semen on the bed and Jane Doe's DNA and think it means abuse," Oral Arg. Tx. at 35. Even if the

experts testify that the government's explanation for how Doe and Johnston's DNA came to be on the comforter is not any more likely than the defense's alternate hypotheses, the jury is likely to wonder why it heard one (or more) full days of testimony about DNA evidence if the data says nothing about the significant factual disputes that the jury needs to resolve. In a myriad of ways, then, such testimony would only invite speculation and confusion.

Second, separate and apart from the disputed issues of "how and why" DNA from Doe and Johnston came to be in these stains, the record provides substantial cause for concern about whether a jury will understand the limitations on Green's ability to "identify" the individuals whose DNA is on the comforter. As previously noted, Green and Kalafut were clear about the fact that the likelihood ratios Green generated from these mixed DNA samples merely test the likelihood of "competing propositions"; they do not establish that Doe or Johnston are, in fact, donors to the mixtures, or that the DNA "matches" their own. *See, e.g.,* Kalafut Tx. at 75 ("Now, we can't say that Jane Doe is the contributor. What we do is an evaluation that says the evidence makes more sense, it has a higher likelihood, if she is a contributor than if she is not"); Green Tx. at 83 ("The likelihood ratio . . . takes the two competing ideas that I set up and says which one is more likely, given this data."). There is some risk that a jury will fail to grasp that distinction in any case involving probabilistic genotyping. But here, with a complex data set involving three-, four-, and five-person mixtures from a shared item in a six-person household, there is a far greater risk that

the jury will hear "1 in 1.0 quintillion" and be extremely confused about its context and meaning.

The government now appears to recognize that distinction. *See* Oral Arg. Tx. at 11 (discussing "relative probability of those contributors being contributors to those mixtures versus not," and clarifying that Green's data is evidence from which a jury could "infer[]" that Doe's and Johnston's DNA is present, but do not prove those assertions). Yet in its pre-hearing opposition brief, the government repeatedly – and incorrectly – characterized Green's testing as conclusive evidence that DNA from Johnston, Monica, and Doe *was present* in these mixed stains. *See, e.g.,* Gov. Opp. at 56–57 (stating that five stains tested by Green "contained mixtures originating from three contributors – *which profiles matched the defendant, Monica, and Doe*") (emphasis supplied); *id.* at 62 (stating that Green made a "finding that *the defendant's and Doe's DNA were contained* in six of the seven semen stains" (emphasis supplied)); *id.* at 68 (stating that "in each of the six samples [tested by Green], *distinct DNA profiles could be determined* for the defendant, Monica, and Doe" (emphasis supplied)). But of course, Green could not – and did not – make any such findings from these complicated mixtures. Nor did she do so from the STRmix likelihood ratios that she calculated.

The Court again underscores that the record does not in any way suggest that the government intentionally distorted Green's conclusions when it filed its brief in opposition to the motion. But it is troubling that even after the government not only had Green's August 2021 Report of Analysis and underlying data in hand for over

three years, but also had ample time to pore over her case file and confer with Green, Kalafut, and others about her findings after the defense filed its motion to preclude, the government still failed to grasp some of the most basic limitations of Green's DNA analysis. While by no means the only (or even a principal) ground for precluding Green's testimony, this record only adds to the Court's existing concerns that the risk of confusion and misleading the jury vastly exceeds whatever limited probative value the DNA evidence might otherwise provide.

CONCLUSION

For the foregoing reasons, the Court grants Johnston's motion to preclude the government from offering expert testimony regarding the DNA evidence generated by USACIL in this case.

In light of this ruling, the Court does not reach that portion of the defense's motion that sought to preclude the government from offering Green's testimony as to her serology results, *i.e.,* her conclusion that seminal fluid was present on certain stained areas of the Johnstons' bedspread. This issue may now be moot. However, if either party wishes to offer evidence about Green's testing for seminal fluid (whether positive or negative) on the various items she received, they shall inquire of the opposing party whether there is any remaining objection to such testimony, and so advise the Court.

SO ORDERED.

/s/ Nina R. Morrison
NINA R. MORRISON
United States District Judge

Dated: March 30, 2025
Brooklyn, New York